

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:56:23 ; Search time 37.5 seconds
(without alignments)
10481.029 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattccattgtgtggga.....tgggatcctaaaaa 2043

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09645078/runat_06052004_104600_21755/app_query.fasta_1.2183
-DB=PIR_78 -QMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USRF=US09645078.cgn_1.1.52/runat_06052004_104600_21755 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	628.5	16.8	484	2 JC7350	N-acetylglucosamin
2	615.5	16.4	486	2 JC7351	N-acetylglucosamin
3	598.5	15.7	484	2 JE0261	N-acetylglucosamin
4	500.5	13.4	458	2 A57397	chondroitin 6-sulf
5	168	4.5	307	2 E95934	probable enzyme, C
6	147	4.0	660	1 Q0B23	BHLFI protein - hu
7	147	4.0	1433	2 S21826	collagen alpha 1(I
8	145	3.9	1464	1 CGHUL5	collagen alpha 1(I
9	143.5	3.8	1172	1 TSHU12	thrombospondin 2 p
10	140.5	3.8	1464	2 S59856	collagen alpha 1(I
11	138.5	3.7	940	2 JE0291	FB19 protein - hum
12	138.5	3.7	1373	1 A43291	collagen alpha 2(I
13	138.5	3.7	1487	1 CGHU6C	collagen alpha 1(I
14	138	3.7	319	2 F75420	hypothetical prote

C 15	138	3.7	671	1 CGRTL5	collagen alpha 1(I
C 16	137.5	3.7	1496	1 CGHU2V	collagen alpha 2(V
C 17	136	3.7	1049	1 CGBO7S	collagen alpha 1(I
C 18	136	3.7	1964	2 T09059	notch4 - mouse
C 19	135	3.6	784	2 JQ0317	hypothetical 82x p
C 20	134.5	3.6	1418	2 T45467	collagen alpha 1(I
C 21	134	3.6	337	2 T23794	hypothetical prote
C 22	134	3.6	1042	1 CGCH1S	collagen alpha 1(I
C 23	133.5	3.6	1419	2 A41182	collagen alpha 1(I
C 24	133.5	3.6	1487	2 B41182	collagen alpha 1(I
C 25	133	3.6	825	1 EDBEXD	immediate-early pr
C 26	133	3.6	1215	2 T32734	myosin-1A - Acanth
C 27	133	3.6	1360	2 T33922	hypothetical prote
C 28	132	3.5	286	2 S34665	collagen, cuticula
C 29	132	3.5	325	2 T32248	hypothetical prote
C 30	131	3.5	636	2 S41067	collagen alpha 1(I
C 31	131	3.5	2944	2 A54849	collagen alpha 1(V
C 32	130.5	3.5	316	2 T19288	hypothetical prote
C 33	130.5	3.5	1122	2 T14180	collagen alpha 2(I
C 34	130.5	3.5	1700	2 S08167	LRG5 protein - Chl
C 35	129.5	3.5	1414	1 S23809	collagen alpha 2(I
C 36	128	3.5	632	2 S42731	collagen alpha 1 c
C 37	127.5	3.4	1184	2 S50832	atrophin-1 - human
C 38	127	3.4	1106	2 JQ0405	hypothetical 119.5
C 39	126	3.4	1366	1 CGHU2S	collagen alpha 2(I
C 40	126	3.4	1712	1 CGHU2B	collagen alpha 2(I
C 41	125.5	3.4	316	2 T19291	hypothetical prote
C 42	125.5	3.4	640	2 T08179	collagen alpha 1(X
C 43	125	3.4	680	1 CGHU1D	collagen alpha 1(I
C 44	125	3.4	1466	1 CGHU7L	collagen - sea urc
C 45	124.5	3.4	290	2 A32249	

ALIGNMENTS

RESULT 1

JC7350 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C:Species: Mus musculus (house mouse)

C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000

C:Accession: JC7350

R:Uchimura, K.; Pasakany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, N.;

Biochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a

A:Reference number: JC7350; MUID:20374462; PMID:10913333

A:Accession: JC7350

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB040710

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylati

sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in

C:Keywords: glycolysis; sulfate transport; sulfotransferase; sulfotransferase protein

Alignment Scores:				
Pred. No.:	2.22e-39	Length:	484	
Score:	628.50	Matches:	154	
Percent Similarity:	48.28%	Conservative:	71	
Best Local Similarity:	33.05%	Mismatches:	177	
Query Match:	16.80%	Indels:	64	
DB:	2	Gaps:	11	

US-09-645-078-1 (1-2043) x JC7350 (1-484)

QY	19	TACGGAACACACAGAGAGGGTAGAGGAGAGAAAGCGCATGGCCCGG---CTAGCAGTG---72	
Db	35	HisserGluGlnAspLysGlyArgAsnCysProGlyLeuGlnArgSerLeuGlyValTrp54	
QY	73	AGCTCTCAAAGCAGCAGGAGCCCAAGCCCAAGGCTCTTCACATTCAGCACAAATGCT132	
Db	55	SerLeuGluAlaAlaAlaAlaGlyGluArgGluGlnGlyAlaGluValArgSerLeuAla74	
QY	133	ACTGCCTAABAAATGAAGCTCCTGCTCTTCTCGTTTCCAGATGCCCATCTTGGTCT192	

Qy	481	GAGACAG-----TCGAGCCTCTTTCAGTCGGGGAAACAG	513
Db	178	aAlaArgAlaProAspThrAlaAenLeuThrAlaAlaLeuPheArgTrpArthrAs	198
Qy	514	CGGGCCCTGTGTTCTGCACCTGCCTGTGACATCATCCCAAGAATCAA-----	562
Db	198	nLysValIleCysSerProProLeuCysProGlyAlaProArgAlaArgAlaGluValGl	218
Qy	563	-ATCATCCCCGGGCTCACTGCAGGCTCCTCTGCAGTCAACAGGCCCTTTGAGGTGGTGA	621
Db	218	yLeuValGluAspThrAlaCysGluArgSerCysProProValAlaIleArgAlaLeuGl	238
Qy	622	GAAGGCTGGCGCTCTCACGCCACGTGGTGCTCAAGSAGGTGCCTTCTTCAACCTGCA	681
Db	238	uAlaGluCysArgIlysTyPrProValValIleLysAspValArgLeuLeuAspLeuGl	258
Qy	682	GTCCCTCTACCCTGTGTGAAGACCCCTCCCTCAACTGCATCTCGTGCACCTGGTCCG	741
Db	258	yValLeuValProLeuLeuArgAppProGlyLeuAsnLeuLysValValGlnLeuPheAr	278
Qy	742	GGACCCCGGGCCGTGTTCGTTCCCGAGAACGCACAAGGAGATCTCATGATTGACAG	801
Db	278	gASPProArgAlaValHisAenSerArgLeuLysSerArgGlnGlyLeuLeuArgGluSe	298
Qy	802	T-----CGCATTGTGTATG-----	814
Db	298	rIleGlnValLeuArgThrArgGlnArgGlyAspArgPheHisArgValLeuLeuAlaHi	318
Qy	815	-----GGGAGCATGAGCAAAAATCAAGAAGGAGGACCAACCCCTA	855
Db	318	sGlyValGlyAlaArgProGlyGlyGlnSerArgAlaLeuProAlaAlaProArgAlaAs	338
Qy	856	CTAT-----GTGATCGAGTCTATCTGCGCAAGCCAGCTGGAGATCTCAAGAC	903
Db	338	pPhePheLeuThrGlyAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuLeuPh	358
Qy	904	CATCAGTCTTGGCCAGGCCCTGCAGGAACGCTACCTGCTTGTGGCTATGAGGACCT	963
Db	358	eAlaArgGlyAlaProAlaTrpLeuArgArgArgTyrrLeuArgLeuArgTyrrGluAspLe	378
Qy	964	GGCTCGAGCCCTGTGGCCACAGACTCCCGAATGTATGAAATCGTGGGATTTGGAAATCTT	1023
Db	378	uValArgGlnProArgAlaGlnLeuArgArgLeuLeuArgPheSerGlyLeuArgAlaLe	398
Qy	1024	GCCCCATCTTCAGACCTGGGTGCATAACATCACCGAGCGCAAGGCATGGGT---GACCA	1080
Db	398	uAlaAlaLeuAspAlaPheAlaLeuAsnMetThrArgGlyAlaAlaTyrrGlyAlaAspAr	418
Qy	1081	CGCTTTCACAAATGCCAGGATGCCCTTAATGTCTCCCGAGGCTTGGCGCTGGTCTTT	1140
Db	418	gpProPheHisLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrpArgGluArgLe	438
Qy	1141	GCCCTATCAAAAGTTTTCTCGACITCAGAAAGCCTGTGGCGATGCCATGAATTTGCTGGG	1200
Db	438	uSerArgGluGlnValArgGlnValGluAlaAlaCysAlaProAlaMetArgLeuLeuAl	458
Qy	1201	CTACCGCCACGTCAGATCTGAACAAGAA	1228
Db	458	aTyrr-----ProArgSerGlyGluGlu	465

A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DDBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfate from 3'-phosphoadenoc
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

```
QY 1004 TTCGTGGGATGGAATCTTGGCCCATCTTACAGACCTGGGTGCATACATCACCAGGC 1063
Db 396 PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly 415
QY 1064 AAGGCGATGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCCTCCAG 1123
Db 416 SerGlySerSerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn 435
QY 1124 GCTTGGCGCTGCTTGGCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGGCGAT 1183
Db 436 AlaTrpArgThrAlaLeuThrPheGlnGlnLeuValGluGluPheCysTyrGln 455
QY 1184 GCCATGAATTTGCTGGGTACCGCCAGCTCAGATCTCAACAGACAGAAACCTGTTG 1243
Db 456 ProMetAlaValLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer 475
QY 1244 CTGGATCTTCG 1255
Db 476 LysThrLeuLeu 479

RESULT 4
A57397
chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
C:Accession: A57397
R:Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi,
J. Biol. Chem. 270, 18575-18580, 1995
A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransf
A:Reference number: A57397; MUID:95355490; PMID:7629189
A:Accession: A57397
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-458 <FUK>
A:Cross-references: GB:D49915; MID:g971262; PIDN:BA08655.1; PID:g971263
A:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Alignment Scores:
Pred. No.: 1,03e-29 Length: 458
Score: 500.50 Matches: 120
Percent Similarity: 54.0% Conservatives: 75
Best Local Similarity: 33.24% Mismatches: 133
Query Match: 13.38% Indels: 33
DB: 2 Gaps: 13

US-09-645-078-1 (1-2043) x A57397 (1-458)

QY 227 CTGCTATGAGGACAGCCGCGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 286
Db 104 LeuGlyIleAlaAlaProGluProArgArgHisValLeuLeuMetAlaThrArgThr 123
QY 287 GCGCTCTCTTTGTTGGGCGAGCTTTTGGGCGAGCACCCAGATGTTTCTACCTGATGGAG 346
Db 124 GlySerSerPheValGlyGluPhePheAsnGlnGlnGlyAsnIlePheTyrLeuPheGlu 143
QY 347 CCGCGCTGCGACGTG-----TGGATGACCTTCAGACAGACCGCTGATGCTGCAC 400
Db 144 ProLeuTrpHisIleGluArgThrValThrPheGluProGlyGlyAlaAsnAlaValGly 163
QY 401 ATGGCTGTG-----CGGATCTGATACGGGCGCTCTTGTGCGCATGAGCGTCTTT 454
Db 164 SerAlaLeuValTyrArgAspValLeuGlnLeuLeuLeuLeuCysAspLeuTyrIleLeu 183
QY 455 GATGCTACATGAACCTGGTCCCGGAGACAG-----TCAGCCTCTTTCAGTGGGAG 508
Db 184 GluSerPheIleSerProAlaProGluGluHisLeuThrAlaAlaLeuPheArgArgGly 203
QY 509 AACAGCGCGCGCTGTGTTCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 565
Db 204 SerSerHisSerLeuCysGluGluProValCys-----ThrProSerLeuLysVal 221
QY 566 ATCCCGCGGCTCACTGCGAG---CTCCTGTGCACTCAACAGCCCTTTGAGGTGGAG 622
```

```
Db 222 PheGluLysTyrHisCysLysAsnArgCysGlyProLeuAsnIleThrLeuAlaAla 241
QY 623 AAGGCTCGCGCTCTACAGCCAGCTGTGTCTCAAGGAGGTGCGCTTCTTCAACCTGCAG 682
Db 242 GluAlaCysArgArgLysGlnHisMetAlaLeuLysThrValArgIleArgGlnLeuGlu 261
QY 683 TCCCTCTACCGCTGCTGAAAGACCCCTCCCTCAACCTGCATATCGTCACCTGTCGG 742
Db 262 PheLeuGlnProLeuAlaGluAspProArgLeuAspLeuArgIleIleGlnLeuValArg 281
QY 743 GACCCCGCGCGCTGCTTCCGTTCCGAGAACGCCAACAAAGGAGATCTCATGATTGACAGT 802
Db 282 AspProArgAlaValLeuValSerArg-----MetVal----- 292
QY 803 CCGATGTGATGGGCGACCATGAGCAAAACCTCAAG----- 838
Db 293 ---AlaPheSerGlyLysTyrGluSerTrpLysLysTrpAlaAlaGluGlyGluAlaPro 311
QY 839 ---AAGGAGGACCAACCTTACTATGTGATGAGCTCATCTGCCAAAGC---CAGCTGGAG 892
Db 312 LeuGlnGluAspGlu---ValGlnArgLeuArgGlyAsnCysGluSerIleArgLeuSer 330
QY 893 ATCTACAAACATCCATCCTTGGCCAAAGCCCTGCAGAACCGTACCTGCTGTGGCGC 952
Db 331 AlaGluLeuGlyLeuArgGln---ProArgTrpLeuArgGlyArgTyrMetLeuValArg 349
QY 953 TATGAGGACCTGGCTGCGAGCCCTGTCGCCAGACTTCCCGAATGTATGAATTCGTGGGA 1012
Db 350 TyrGluAspValAlaAlaGalaProLeuArgLysAlaLeuGluMetTyrArgPheAlaGly 369
QY 1013 TTGGAATCTTCCCTCATCTTCCAGACCTGGGTGCATACATCACCAGGAGGAGGCATG 1072
Db 370 IleHisProThrProGlnValGluGluTrpIleArgAlaAsnThrGlnAlaPro---Gln 388
QY 1073 GGTGACCGCTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGC 1132
Db 389 AspSerAsnGlyIleTyrSerThrGlnLysAsnSerSerGluGlnPheGluLysTrpArg 408
QY 1133 TGTCTTTTCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGGCGATGCCATGAT 1192
Db 409 PheSerIleProPheLysLeuAlaGlnValGlnAspAlaCysGluProAlaMetArg 428
QY 1193 TTGCTGGGTACCGCCAGCTCAGATCTGACAGACAGACAGAAACCTGTTGCTGATCTT 1252
Db 429 LeuPheGlyTyrLysLeuAlaSerAlaGlnGluLeuThrAsnArgSerLeuSerLeu 448
QY 1253 CTG 1255
Db 449 Leu 449

RESULT 5
E95934
probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sino
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95934
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95934
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <KUR>
A:Cross-references: GB:AL591985; PIDN:CA49141.1; PID:gi15140625; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
```


Db	333	SerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProAlaAlaAa	352
Qy	875	AGATGACCT-----GCATCACAT	858
Db	353	ArgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArg	372
Qy	857	AGTAGGGTTGGTCTCTCT-----TCCTGAGTTTGGTCTCATGCTGCC	816
Db	373	CysProAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArg	392
Qy	815	CCATCAACATCGCACTGCTCAATCATGAGATCTCCCTTTGGTGGTTCTC-----	768
Db	393	ProProGlyCysPro-----ArgSerAlaArgAsnProGlyCysProArgThrTrpArgArg	411
Qy	767	-----GGGAACGGAACACACGGCCCGGGGGTCCC	741
Db	412	ArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlnArgProSerGlyPro	431
Qy	740	GGACCAAGTGCACATATGCAGTTTGAGGGGGTCTTTCCAGCAGCG-----GGT	690
Db	432	ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly	451
Qy	689	AGAGGACTGCAGGTGGAAGAAGCGCACCTCCTTGGAGCACCACTGGCTGTAGAGAGGGC	630
Db	452	GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGly-----SerGly	469
Qy	629	AGGCCTTCTCCA-----CCACTCAAGG-----	606
Db	470	ProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuPro	489
Qy	605	-----GCTGTTTGACTGC-----ACAGGAGGCC	585
Db	490	GlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAla	509
Qy	584	TGCAGTGAG-----CCCGGGGATGATTTCATCTTGGGTGATGATGT	543
Db	510	AlaAlaGlnArgThrHisArgArgProProGly-----Cys	521
Qy	542	CACAGGCAGGTGCAGAACACAGGGCCCGCTGTTCTCCCACT-----GAAAGAGCGTGG	489
Db	522	ProArgSerAlaArgAsn-----ProGlyCysProArgThrTrpArgArgSerGly	539
Qy	488	ACTGTCTCCGGGACCAGGTTCATGTAGGCATCAAGACCGTCTATGTGCGCACAGAAGA	429
Db	540	AlaGlnArgGlyHisProProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGly	559
Qy	428	CGGCCCGTATCAGATCCCGCACAGGCATGTGCAGCATCCAGCGGTGCTCTGCTGAAGG	369
Db	560	ArgProAlaAla-----ProGlyAlaPro-----	567
Qy	368	TCATCCACAGTGCACAGGGGGGCTCCATCAGTAGAACAACATCTGGGTCTGCC-----	315
Db	568	---GlyThrProAlaAlaProGlyProGlyGly-----GlyAlaAlaValPro	582
Qy	314	CAAAAAGTGCCTCCCAAAAGAACAGACGAGCGCGGAGAACAGACACACAGCGTGA	255
Db	583	SerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProPro-----Ala	600
Qy	254	TGCGTCTGGGTGTGCTTCATAGACAGGAGCTGATGTTGGCTGTACATCTGGGAAGA	195
Db	601	AlaAlaArgLeuProProGlu-----AspLeuAlaAlaAla-----	608
Qy	194	ATAGACCAAGATGGCCATCTGGGAACCCAGAAACAGACAGAGCTTCATTTTTTAGGCA	135
Db	609	GlnGluProArgLeuProGln-----	620
Qy	134	GTAGCATTTGCTGAAGTGAAGACCTTTGGCTTGGGCTTCCCTGCTCTTTTGAGAGG	75
Db	621	-----GlnArgCysProAlaGlyProProProThrArgSerGly	633
Qy	74	CTCACTGCTAGCGGGCCCATGGCTTTTCTCT-----CTACCC	36
Db	634	AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnPro	653

QY 35 TTCTGTGCTTCCTCGTACCCACAACTGG 6
|||
|||
Db 654 GlyCys-----ProArgThrTrp 659

RESULT 7
S21626
collagen alpha 1(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 13-Aug-1999
C:Accession: S57243; S16374; A23982; I49557; S39789; I48300; S21626
R:Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A:Reference number: S57243
A:Accession: S57243
A:Molecule type: mRNA
A:Residues: 1-1453
A:Cross-references: EMBL:J08020; NID:G470673; PIDN:AAA88912.1; PID:G470674
R:Metzgerant, M.; Toman, D.; de Crombrughe, B.; Vucirio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MET>
A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485
R:French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A:Reference number: A23982; MUID:86137403; PMID:3841523
A:Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <FRE>
A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262
R:Monson J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A:Reference number: I49559; MUID:83141374; PMID:8298597
A:Accession: I49559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264
R:Harbers, K.; Kuehn, M.; Dellus, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A:Reference number: I49557; MUID:84170331; PMID:6324198
A:Accession: I49557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <RE>
A:Cross-references: GB:X01698; NID:G192246; PIDN:AAA37330.1; PID:G553881
R:Fenton, S.P.; Lamande, S.R.; Hamaguchi, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8268229
A:Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1
R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect
A:Reference number: I48300; MUID:94344105; PMID:8065328
A:Accession: I48300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>
A:Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487
C:Genetics:
A:Gene: COL1A1
A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A/Title: Regulatory elements in the first intron contribute to transcriptional control c
A/Reference number: A39943; MUID:88097389; PMID:3480516
A/Accession: A39943
A/Molecule type: DNA
A/Residues: 1-34 <SOR>
A/Cross-references: GB:J03559; NID:9180876; PIDN:AAA52052.1; PID:9553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A/Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A/Reference number: I55237; MUID:85130970; PMID:2857713
A/Accession: I55237
A/Status: translation not shown; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-34 <CH2>
A/Cross-references: GB:M10627; NID:9180383; PIDN:AAA51992.1; PID:9553226
R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A/Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
rome, type VII.
A/Reference number: A35233; MUID:90202908; PMID:2318855
A/Accession: A35233
A/Molecule type: protein
A/Residues: 33-52 <WR>
A/Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R:Weil, D.; d'Allesio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A/Title: A base substitution in the exon of a collagen gene causes alternative splicing
A/Reference number: S09400; MUID:89356643; PMID:2767050
A/Accession: S09400
A/Molecule type: mRNA
A/Residues: 156-183 <WBI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4698-4706, 1970
A/Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A/Reference number: A90567; MUID:71038625; PMID:5529814
A/Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
A/Accession: B90567
A/Molecule type: protein
A/Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A/Experimental source: skin
A/Note: evidence for 170-alanine
R:Baetge, B.; Norbom, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F
Eur. J. Biochem. 192, 153-159, 1990
A/Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A/Reference number: S11372; MUID:90382436; PMID:2169412
A/Accession: S11372
A/Molecule type: protein
A/Residues: 175-187, 274-287, 'P', 289 <BAR>
A/Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantino, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A/Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
cooperative melting of intact type I collagen.
A/Reference number: I55342; MUID:92042092; PMID:1718984
A/Accession: I55342
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 258-268; 1347-1357 <DEA>
A/Cross-references: GB:S67495; NID:9239007; PIDN:AA20350.1; PID:9239008
A/Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A/Title: Comparative study of glycoproteins derived from selected vertebrate collagens.
A/Reference number: A92069; MUID:71001508; PMID:4319110
A/Accession: A92069
A/Molecule type: protein
A/Residues: 263-268 <MOR>
A/Experimental source: skin
A/Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A/Title: Segmental amplification of the entire helical and telopeptide regions of the cd

A/Reference number: S15989; MUID:90326017; PMID:2374517
A/Accession: S15989
A/Molecule type: mRNA
A/Residues: 381-302, 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <LAB>
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A/Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of ty
A/Reference number: I52905; MUID:93339042; PMID:8339541
A/Accession: I52905
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 342-352, 'C', 354-359 <W12>
A/Cross-references: GB:S64717; NID:9408195; PIDN:AA27677.1; PID:9408196
A/Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A/Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha
A/Reference number: A90476; MUID:84080385; PMID:6689127
A/Accession: A90476
A/Molecule type: mRNA
A/Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A/Cross-references: GB:K01228; NID:9180391; PIDN:AAA51995.1; PID:9180392
A/Note: sequence partially completed for missing nucleotides by A29439
R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A/Title: Multitoxon deletion in an osteogenesis imperfecta variant with increased type II
A/Reference number: A22161; MUID:85104934; PMID:2381843
A/Accession: A22161
A/Molecule type: DNA
A/Residues: 472-594, 'R', 596-607 <CH3>
A/Cross-references: GB:K03178; GB:K03179; NID:9179612; NID:9179613; PIDN:AAA51847.1; PID
A/Note: the authors translated the codon CGT for residue 595 as Pro
R:Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A/Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
A/Reference number: A35336; MUID:90252792; PMID:2339700
A/Accession: A35336
A/Molecule type: mRNA
A/Residues: 710-720, 'E', 722-737, 'B', 739-745 <WAL>
A/Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R:Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A/Title: Severe type III osteogenesis imperfecta due to glycine substitutions in the c
A/Reference number: I54365; MUID:95187161; PMID:7881420
A/Accession: I54365
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 746-766, 'S', 768-781 <FOR>
A/Cross-references: GB:47867; NID:91009093; PIDN:AA59576.1; PID:91009094
R:Chessier, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A/Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A/Reference number: A47426; MUID:93352646; PMID:8349697
A/Accession: A47426
A/Molecule type: mRNA
A/Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
A/Cross-references: GB:S64596; NID:9407589; PIDN:AA27856.1; PID:9407590
A/Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIPI:136445)
A/Note: does not represent an experimentally determined sequence but three different mut
A/Molecule type: mRNA
A/Residues: 1179-1464 <CH4>
A/Experimental source: normal dermal fibroblast culture
A/Accession: C47426
A/Molecule type: mRNA
A/Residues: 1179-1276, 'H', 1278-1464 <CH5>
A/Experimental source: fetal cell 86-237
A/Accession: D47426
A/Molecule type: mRNA
A/Residues: 1179-1336, 1339-1464 <CH6>
A/Experimental source: fetal cell 86-146
A/Accession: E47426

Db 333 ProProGlyProThrGlyProAlaGly-----ProProGlyPheProGly-Al 348
 QY 641 TGTAGGAGCGGCGAGCCCTTCCACCACCTCAAAGGGCTGTGTAGTCACAGGAGCCCTGC 582
 Db 348 aValGlyAla-----LysGlyGluAlaGlyProGlnGlyProAr 361
 QY 581 AGTGAGCCCGGGGATGATTTCATCTTGTTGGATGATGTACACAGCAGCGTGCAGAACACA 522
 Db 361 gGlySerGluGlyPro-----GlnGlyValArgGlyGluProGlyProPr 376
 QY 521 GGGCCCGGCTGTTCTCCCACTGAAGAGGCTGGACTG-----TCTCGGGGACCAAGGTT 468
 Db 376 oGlyProAla-----GlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspG 392
 QY 467 CCATGTAGGCATCAAAGACGCTCATGTCCGACACAGAGAGCGCCCGTATCAGATCCCGCA 408
 Db 392 yGlnProGlyAlaLysGlyAlaAsnGlyAla----- 402
 QY 407 CAGCCATGTGCAGCATCCAGC-----GGTGTCTCTCTTTGA 372
 Db 403 -----ProGlyIleAlaGlyAlaProGlyPheProGlyAlaArgGlyPr 417
 QY 371 AGGTCAATCCACAGTCGCGGGGCTTCATCAGGTAGAAACATCTGGTGTCTGCCCAA 312
 Db 417 oSerGlyProGlnGlyProGlyGly--ProProGlyProLysGly---AsnSerGlyGlu 435
 QY 311 AAAGCTGCCCCACAAAAGAGACGACGAGCCCGAGGAACACAGACCACTGCATGC 252
 Db 436 ProGlyAlaProGlySerLysGlyAspThrGlyAlaLysGlyGluProGlyProValGly 455
 QY 251 GCTCGGGCTGTGCTTCATAGACAGGAGCTGATGTTGTGGCTGTACATGTGGAAAGATA 192
 Db 456 ValGlnGlyProProGlyProAlaGlyGlu-----GluGlyLysArgGlyAlaArgGly 473
 QY 191 GAGCCAAGATGGCCATCTGGGAAACACAGAAACAGCAGGAGCTTCATTTTTTTAGCAGTA 132
 Db 474 GluPro---GlyProThrGlyLeuProGlyProProGlyGlu----- 486
 QY 131 GCATGTGTCTGAAATGGAAGACCTTGTGGCTTGGGCTTCCTCTGCTGCT 84
 Db 487 -----ArgGlyGlyProGlySerArgGlyPheProGlyAla 498
 RESULT 9
 TSHUP2
 thrombospondin 2 precursor - human
 C;Species: Homo sapiens (man)
 C;Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
 C;Accession: A47379; A42173
 R;LaBell, T.L.; Byers, P.H.
 Genomics 17, 225-229, 1993
 A;Title: Sequence and characterization of the complete human thrombospondin 2 cDNA
 A;Reference number: A47379; MUID:94010892; PMID:8406456
 A;Accession: A47379
 A;Molecule type: mRNA
 A;Residues: 1-1172 <LA8>
 A;Cross-references: NID:G307505; PIDN:AAA03703.1; PID:G307506
 R;LaBell, T.L.; Milewicz, D.J.; Distche, C.M.; Byers, P.H.
 Genomics 12, 421-429, 1992
 A;Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expression
 A;Reference number: A42173; MUID:92217961; PMID:1559694
 A;Accession: A42173
 A;Molecule type: mRNA
 A;Residues: 560-1172 <LA2>
 A;Cross-references: GDB:M81339
 A;Experimental source: fibroblast
 A;Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:95096)
 C;Genetics:
 A;Gene: GDB:THB32; TSP2
 A;Cross-references: GDB:128789; OMIM:188061
 A;Map position: 6q27-6q27
 C;Complex: homotrimer, disulfide linked
 C;Function:
 A;Description: participates in cell migration and adhesion, and in platelet aggregation

C; Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vWf C; Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; trimer F; 1-18/Domain: signal sequence #status predicted <SIG> F; 19-1172/Product: thrombospondin 2 #status predicted <MAT> F; 119-1172/Product: thrombospondin 2 #status predicted <MAT> F; 119-377/Domain: von Willebrand factor type C repeat homology <VWC> F; 380-431/Domain: thrombospondin type 1 repeat homology <THR1> F; 436-492/Domain: thrombospondin type 1 repeat homology <THR2> F; 493-549/Domain: thrombospondin type 1 repeat homology <THR2> F; 553-588/Domain: EGF homology <EGF1> F; 552-691/Domain: EGF homology <EGF> F; 928-930/Region: cell attachment (R-G-D) motif F; 316,330,457,584,710,1069/Binding site: carbohydrate (Asn) #status predicted F; 167-226/Disulfide bonds: #status predicted F; 266-270/Disulfide bonds: interchain #status predicted F; 612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Alignment Scores:
Pred. No.: 0.0121 Length: 1172
Score: 143.50 Matches: 131
Percent Similarity: 29.09% Conservative: 61
Best Local Similarity: 19.85% Mismatches: 229
Query Match: 3.83% Indels: 239
DB: 1 Gaps: 31

US-09-645-078-1 (1-2043) x TSHUP2 (1-1172)

Qy	199	CCACATGTACAGCCACACATCATCGCTCTCTATGAGGCACAGCCCGAG-----	250
Db	248	ProHisValThrThrGluTyrValGlyPro-SerSerGluArgArgProGluValCysG1	267
Qy	251	-----CCCATGCACGTCGTGGT	267
Db	267	uArgSerCysGluGluLeuGlyAsnMetValGlnGlnLeuSerGly-LeuHisValLeuVa	287
Qy	268	T-----CTGTCTCTCTGGCGTCTGGCTCTCTTTTGTGGGGCAGCT	309
Db	287	laAsnGlnLeuSerGluAsnLeuLysArgValSerAsnAspAsnGlnPheLeuTyrGluLe	307
Qy	310	TTTTGGGCAGCACCACAGATGTTTCTACTCATGTGAGGCCCGCTGGCAGCTGTGGATGAC	369
Db	307	uileGlyGlyProProLysThr--ArgAsnMetSerAlaCysTrpGlnAspGlyArgph	326
Qy	370	CTTCAGCAGAGCAGCCGCTGGATGCTCACAATGGCTGTGGGATCTGATACGGGCGGT	429
Db	326	ePheAlaGluAsnGluThrTrp-----ValValAspSerCys	338
Qy	430	CTTCTTGTGGCAGCATGAGCGCTTTTGATCCCTCATGGAACCTGGTCCCGGAGACATC	489
Db	338	sThrThrCysThrCysLysLysPhe-----	346
Qy	490	CAGCCTCTTTCAGTGGGAGAACAGCGGGCCCTGTGT-----TCTGCACCTGCGCTG	540
Db	347	-----LysThrTrpCysHisGlnIleThrCysPro-ProA	358
Qy	541	TGACATCATCCACAAAGATGAATCATCCCCGGGCTCACTGCAGGCTCCTGT-----	593
Db	358	laThrCys-----AlaSerProSerPheValGluGlyCysCysProS	373
Qy	594	-----GCAGTCAACAGCCCTTTCAGGTGTGGAGAGGCTCGCGCTCCCTACAG	642
Db	373	erCysLeuHisSerValAspGlyGluGluGlyTrp-----SerProTrrpA	388
Qy	643	CCACGTGG-----TGCTCAAGAGGTGGCGTCTTCACTGCAGTCCCTCTACCCCGCT	696
Db	388	laGluTrpThrGlnCysSerValThrCysGlySerGlyThrGlnGlnArgGlyArgSerC	408
Qy	697	GCTGAAGAGCCCTCCCTCAACCTGCATA-----TGCTGCACC-	734
Db	408	ysAspValThr---SerAsnThrCysLeuGlyProSerIleGlnThrArgAlaCysSerL	427
Qy	735	-----TGCT	738
Db	427	eusSerIvsCysAspThrArgGlnAspGlyGlyTrpSerHisTrrpSerProTrrpS	447

QY 284 AGCCGACGAGACAGACACGACGATGCGATCGGCTC 248
Db 849 uGlyProGlyHisGlyProHisGlyHisArgLeu 861
RESULT 12
collagen alpha 2(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A43291; A54328
R:Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.
Genomics 13, 1345-1346, 1992
A:Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collagen
A:Reference number: A43291; MUID:92372043; PMID:1505972
A:Accession: A43291
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1373 <PHI>
A:Cross-references: GB:X58251; NID:G50488; PIDN:CAA41205.1; PID:950489
A:Note: sequence extracted from NCBI backbone (NCBI:P112027)
R:Phillips, C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.
J. Invest. Dermatol. 97, 980-984, 1991
A:Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polymera
A:Reference number: A54328; MUID:92084969; PMID:1748823
A:Accession: A54328
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-110 <PH2>
C:Genetics:
A:Gene: COL1A2
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
P:1145-1373/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Alignment Scores:
Pred. No.: 0.0298 Length: 1373
Score: 138.50 Matches: 104
Percent Similarity: 30.67% Conservative: 15
Best Local Similarity: 26.80% Mismatches: 134
Query Match: 3.74% Indels: 136
Gaps: 24
DB: 1
US-09-645-078-1 (1-2043) x A43291 (1-1373)
QY 1127 AGCCTGGGAGACATTAAGGCGATCCCTGGCATTCCTGGAAAGCGTGTGTCACCATGC 1068
Db 512 LysProGlyGluArg---GlyHisPro---GlyLeuAlaGlyAla---GlyAlaProGly 529
QY 1067 CCTTCCTCGGTGATGTTATGACCCAGGTCTGAAGTAGTGGGCAAGAATTCATCCCA 1008
Db 530 Pro----- 530
QY 1007 CGAATTCATACATTCGGAAGTCTCGGCCACAGGGCTCGAGCCAGTCTCATAGCGCA 948
Db 531 -----AspGlyAsnAsnGlyAlaGlnGly---ProProGlyProGlnGlyVal 545
QY 947 CAGCAGGTAGCTTCCTCGAGGGCTTGGGCAAGGACTGGATGCTTGTAGATCTCCA 888
Db 546 GlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 557
QY 887 GCTGGCTTTGGCAGATGACCT-----GCATCATATAGTAGGTGGTCTCTCTCT 837
Db 558 ProGlyPheGlnGlyLeuProGlyProSerGlyThrThrGlyGluValGlyLysProGly 577
QY 836 TGAGTTTTGCTCATGCTGCCCATCACAATCGCAATCGCAATCATGATGATCTCCCTTTG 777
Db 578 -----GluArgGlyLeuPro--- 582
QY 776 TGGCTTCTCGGGAACGGAACACGCGCGGGGTCCCGACACAGGTGCGACGATATGAGGT 717
Db 583 -----GlyGluPheGlyLeuProGlyProAlaGlyPro----- 593
QY 716 TGAGGAGGGGTCTTTCAGCAGCGGTAGAGGACTGACAGGTGTAAGA---AGCGCACT 660

A:Molecule type: mRNA
A:Residues: 1-940 <TOT>
A:Cross-references: GB:Y13247; NID:92117159; PIDN:CAA73697.1; PID:92117159
C:Genetics:
A:Gene: FB19
A:Map position: 6p21.3
Alignment Scores:
Pred. No.: 0.0275 Length: 940
Score: 138.50 Matches: 77
Percent Similarity: 31.06% Conservative: 14
Best Local Similarity: 26.28% Mismatches: 94
Query Match: 3.74% Indels: 108
Gaps: 16
DB: 2
US-09-645-078-1 (1-2043) x JE0291 (1-940)
QY 992 CGGAATCTGGGACACAGGGCTCGAG-----CCA 963
Db 633 GlyProGlyProGlyProGlyMetGlnHisPheProProGlyProGlyProMetPro 652
QY 962 GGTCTCTATAGCGCACAGCAGTAGCGTTCCTGACGGCCCTTGGCAAGGACTGGATGG 903
Db 653 GlyProHisGlyGlyProGlyGlyProValGlyPro----- 664
QY 902 TCTTGTAGATCTCAGCTGCTTGGCAGATGACCTGCATCATAGTAGGTGGTCTCT 843
Db 665 -----ArgLeuLeuGlyPro 669
QY 842 CTTCTTGTAGTTTGTCTCATGCTGCCCATCAATGCGACTGTCAATCATGAGATCTC 783
Db 670 Pro-----ProProArgGlyGlyAspProPheTTPAspGly 682
QY 782 CTTTGTGCTTCTCGGACGACGACACGCGCGGGTCCCGACGACGATGACATAT 723
Db 683 ProGlyAspProMetArgGlyGlyProMetArgGlyGlyProGlyProGlyPro----- 700
QY 722 GCAGTTGAGGGGCTTTCAGCGCGGTAGAGGACTGCGAGTTCAAGACGCGCA 663
Db 701 -----GlyProTyHisArgGlyArgGlyGlyArgGlyGlyAsnGluPro 715
QY 662 CCTCTTGTAGCACCA-----CGTGTGTAGAGCGCGAGGCT-----TTCACCA 615
Db 716 ProProProProProPheArgGlyAlaArgGlyGlyArgSerGlyGlyGlyProPro 735
QY 614 CCTCAAGGGCTGTGACTGACAGGAGCTGACAGTGGAGCCCGGGGATGATTCATCTT 555
Db 736 AsnGlyArg-----GlyGlyProGlyGlyGlyMet---Val 746
QY 554 GTGGGATGATGTACAGGAGGTGCAGAACAGCGCCCGCTCTTCT----- 507
Db 747 GlyGlyGlyGlyHisArgProHisGluGlyProGlyGlyMetGlyAsnSerSerGly 766
QY 506 -----CCACTGAA-----AGAGCTGACTGTCTCGGGACCGAGTTCCATGTAG 459
Db 767 HisArgProHisGluGlyProGlyGlyGlyMetGlySerGlyHisArgProHisGluG 786
QY 458 CATCAAGACGCTCATGTGCGCACAGAGAGCGGCCGTATCATGATCCCGCACAGCCATGT 399
Db 786 Y-----ProGlyGlySerMetGlyGlyGlyGlyGlyHisArgProHis----- 800
QY 398 CGAGCATCCAGCGGTCTCTCTGTAAGGTTCATCCACAGTGGCA----- 353
Db 801 -GluGlyProGlyGlyGlySerGlyGlySerGlyHisArgProHisGluGlyProG 820
QY 352 -----GGCGGCTCCATCAGTAGAAGAACATCTGGGTGCTGCCCAAAA 309
Db 820 yGlyGlyMetGlyAlaGlyGlyGlyHisArg----- 830
QY 308 GCTGCCCAAAA-----AGAGAGCCAG 285
Db 831 ----ProHisGluGlyProGlyGlySerMetGlyGlySerGlyGlyHisArgProHisG 849

A:Reference number: I38867; MUID:95150028; PMID:7847372
A:Accession: I38867
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509, 'TIIA'
A:Cross-references: EMBL:U15195; NID:9557053; PIDN:AA60370.1; PID:9557054
R:Ramirez, P.
submitted to the EMBL Data Library, December 1988
A:Reference number: S04892
A:Accession: S04892
A:Molecule type: mRNA
A:Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214, 'RAM'
A:Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050
R:Vikkula, M.; Peltonen, L.
PEBS Lett. 250, 171-174, 1989
A:Title: Structural analyses of the polymorphic area in type II collagen gene.
A:Reference number: S05000; MUID:89325561; PMID:2753125
A:Accession: S05000
A:Molecule type: DNA
A:Residues: 630-640, 'A', 642-785, 'VIK2'
A:Cross-references: EMBL:X16158; NID:929951; PIDN:CAA34278.1; PID:9335018; PIDN:CAA3427
PIDN:CAA34283.1; PID:9335023; PIDN:CAA34284.1; PID:9335024
R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D.
J. Biol. Chem. 267, 22522-22526, 1992
A:Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
A:Reference number: A44309; MUID:93054548; PMID:1429602
A:Accession: A44309
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA; mRNA
A:Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',
A:Cross-references: GB:100977; NID:9180812; PIDN:AB23914.1; PID:9589774
A:Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence we
A:Note: this translation is not annotated and this publication is not cited in GenBank
A:Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R:Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A:Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A:Reference number: S16502; MUID:90251662; PMID:2339128
A:Accession: S16502
A:Molecule type: DNA
A:Residues: 1164-1184, 'GPSKDGANGIRGPI', 1185-1199, 'TII2'
A:Cross-references: EMBL:M37126; NID:9180808; PIDN:AAA52037.1; PID:9180809
A:Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R:Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 92, 2555-2559, 1995
A:Title: Identification and characterization of the human type II collagen gene (COL2A1)
A:Reference number: A02858; MUID:85190534; PMID:3857598
A:Accession: A02858
A:Molecule type: DNA
A:Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487, 'CHE'
A:Cross-references: GB:J00116; NID:9180395; PIDN:AAA51997.1; PID:9180396
R:Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A:Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A:Reference number: A27280; MUID:88067771; PMID:2825137
A:Accession: A27280
A:Molecule type: DNA; mRNA
A:Residues: 1175-1487, 'ELI'
A:Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:930097
A:Experimental source: fetal epiphyseal cartilage
R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A:Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A:Reference number: A57033; MUID:87099927; PMID:3800925
A:Accession: A57033
A:Molecule type: protein
A:Residues: 'XG', 1244-1246, 'N', 1248, 'X', 1250-1255, 1295-1305, 1395-1408, 'VAN'
A:Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal prope
R:Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A:Title: Isolation and characterization of genomic clones corresponding to the human typ
A:Reference number: A21733; MUID:84118798; PMID:6320112
A:Accession: A21733

A:Molecule type: DNA
A:Residues: 1245-1295, 'STR1'
A:Cross-references: EMBL:X00339; EMBL:X00298; NID:9394699; PIDN:CAA25092.1; PID:94378975
A:Accession: B21733
A:Molecule type: DNA
A:Residues: 894-909, 'PE', 'STR2'
A:Cross-references: GB:K01785; NID:930035; PIDN:CAA25082.1; PID:91335032
R:Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A:Title: Isolation and partial characterization of genomic clones coding for a human pro
gene.
A:Reference number: A24561; MUID:86104139; PMID:3002437
A:Accession: A24561
A:Molecule type: DNA
A:Residues: 1296-1358, 'NUN2'
A:Cross-references: GB:M12048; NID:9180017
A:Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A:Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
R:Sangiorzi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A:Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
A:Reference number: I37249; MUID:85215609; PMID:2987845
A:Accession: S59491
A:Molecule type: DNA
A:Residues: 7-28, 'R', 99-114, 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1
A:Accession: I84453
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 7-28, 'S2AN2'
A:Cross-references: GB:M23759; NID:9180845; EMBL:X03320; GB:M24938; NID:930104
A:Note: the GenBank PID is based on an incorrect reading frame
A:Accession: I37250
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 541-560, 'S2AN3'
A:Cross-references: EMBL:X02378; GB:M23870; NID:930107; PIDN:CAA26227.1; PID:9929621
A:Accession: I37251
Alignment Scores:
Pred. No.: 0.0303 Length: 1487
Score: 138.50 Matches: 87
Percent Similarity: 35.47% Conservative: 18
Best Local Similarity: 29.39% Mismatches: 119
Query Match: 3.74% Indels: 72
DB: 1 Gaps: 17
US-09-645-078-1 (1-2043) x CGHU6C (1-1487)
QY 1142 GCAAGACAGCGCC---AAGCGTGGGAGACATTAAAGGGCATCCCTCGCATTTGTGTGA 1086
DB 418 AlaLysGlySerAlaGlyAlaProGly-----IleAlaGly 429
QY 1085 AAGCGTGTCCACCGTGCCTGCGGTGATGTTATGACCCAGGTCTGAAGATGGG 1026
DB 430 AlaProGlyProGlyPro-----ArgGlyProProAspProGlnGly 444
QY 1025 GCAAGAATTCCAATC-----CCACGAATTCATACATTC--- 993
DB 445 AlaThrGlyProLeuGlyProLysGlyGlnThrGlyLysProGlyIleAlaGlyPheLys 464
QY 992 GGAAGTCTGGCCACAGCGGTCTGAGCAGGTCTCTATCGGCACAGCAGGTAGCGTT 933
DB 465 GlyGluGlnGlyProLysGly---GluProGlyProAlaGlyProGlnGlyAlaProGly 483
QY 932 CCGCAGGCGCTTGGGCAAGGACTCGATGCTTGTAGATCTCCAGCTGGCTTGGCAGA 873
DB 484 ProAlaGlyGluGluGlyLysArgGly---AlaArgGlyGluProGlyValGly--- 501
QY 872 TGACCTGCATCACATAGTAGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 825
DB 502 -----ProIleGlyProGlyGluArgGlyAlaProGlyAsnArg 515
QY 824 -----CATGCTGCCCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 798

Db 516 GlyPheProGlyGlnAspGlyLeuAlaGlyProLysGlyAlaProGlyGluArgGlyPro 535
QY 797 CAATCATGATCTCCCTTTGCGTCTCGGACCGAACACCGCCCGGGGTCCCGGA 738
Db 536 SerGlyLeuAlaGlyProLysGlyAlaAsnGlyAspProGlyArgProGlyGluProGly 555
QY 737 ---CCAGGTGACAGATATGCGAGGTTCAGGAGGGGTCTTCAGACGCGGGTAGAGGACT 681
Db 556 LeuProGlyAlaArgGlyLeuThr---GlyArgProGlyAspAlaGlyProGlnGlyLys 574
QY 680 GCAGTTGAAGAGCGGACCTCTTGTGAGCACACCGTGTGTAGAGCGGCGGCTTCT 621
Db 575 ValGlyProSerGlyAlaProGly---GluAspGlyArgProGly 588
QY 620 CCACCA---CCTCAAAGGCGTGTGACTGCACAGGAGCTCAGTGCAGCGCGGGATGA 564
Db 589 ProProGlyProGln---GlyAlaArgGlyGlnProGly--- 600
QY 563 TTTCATCTGTGGGATGATGACACAGGAGGTGAGAGCGGCGGCTGTTCTCC 504
Db 601 ---ValMetGlyPheProGlyProLysGlyAlaAsnGlyGluProGlyLysAlaGly 618
QY 503 ACTGAAGAGGCTGACTGTCTCGGGACCGATATCAGATC---CGCACACCGCATGTC 444
Db 619 Glu-LysGlyLeuProGlyAlaProGlyLeuArgGlyLeuProGly---LysAspGlyG 637
QY 443 TGTGCAACAAGAGCGGCGGATATCAGATC---CGCACACCGCATGTC 396
Db 637 uThrglyAlaGlyProProGlyProAlaGlyProAlaGlyGluArgGlyGluGlnG 657
QY 395 GCATCAGCGGTGCTGCTGTGAGGTTCATCCACAGTGCAGGC 350
Db 657 yAlaProGlyProSerGlyPheGlnGlyLeuProGlyProGly 672

RESULT 14
F75420
hypothetical protein - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C/Accession: F75420
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036996; PMID:10567266
A/Accession: F75420
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-319 <WHI>
A/Cross-references: GB:AE001971; GB:AE000513; NID:G6458972; PIDN:AAF10810.1; PID:G645898
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR1229
A/Map position: 1

Alignment Scores:
Pred. No.: 0.0238 Length: 319
Score: 138.00 Matches: 79
Percent Similarity: 30.82% Conservative: 15
Best Local Similarity: 25.90% Mismatches: 87
Query Match: 3.69% Indels: 124
DB: 2 Gaps: 23

US-09-645-078-1 (1-2043) x F75420 (1-319)

QY 219 TCAGTCCCTGCTCATGAGGACAGCGCCGAGCGCATCGCTGCTGG 266
Db 77 SerAlaProCys-----ProSerAlaAlaGlyCysTrpAsnSerSerAla 91
QY 267 -----TTCTGTCTCTGCGCTCTGGCTCTCTTTTCTGGGCG 305

Db 92 ArgValArgAspArgArgSerCysSerProSerArgArgCys-----TrpGly 108
QY 306 AGCTTTTGGGAGACACCCAGATGTTTCTACCTGATGGAGCCGCTGGCAGGTGGA 365
Db 109 -----TrpProSerLysGly---CysArg 115
QY 366 TGACCTTCAAGCAGACACCGCTGGA 116
Db 116 ---ProGlyLysArgGlyGlyAlaGlyTTPProArgArgSerLysThrTrpCys 134
QY 396 TGCACATGCTGTGGGATCTGATACGGCGCTTCTTGTGCGACATGACGCTCTTTG 455
Db 135 GlyGlyTrpArgAlaAlaSerThrSerGlySerSerCysAlaAlaArgLysArg 154
QY 456 ATGCTTACATGAACCTGCTCCCGAGACACATCCAGCTCTTTCAGTGGGAGAACAG 515
Db 155 ArgArgHisTrp-----GlySerSerGlyProSer---SerGlyCysSerAla 169
QY 516 GGGCCCTGT---GTTCTGCACCTGCTGTGACATCATCCCAAGATGAATCATCCGCC 572
Db 170 AsnThrThrArgAlaArgHisAlaProAla-----ArgSerSerPro 183
QY 573 GGGCTCCTGACGGCTCTGTCAGTCAACAGCCCTTTGAGG---TGGTGGAGAAAG 626
Db 184 LeuAlaSerThrGly-----HisSerProAlaAlaSerGlyTrpGlyArg--- 198
QY 627 CTGCGCTCTCTACACCCAGCTGTGC-----TCAAGGAGGTGCGCTTTCACCTGC 680
Db 199 ---SerProSerArgAsnTrpCysGlyAlaAlaArgProSerValSerSerGlyArg 216
QY 681 AGTCCCTCTACCGCTGCTGAAAGACC-----CCTCCCTCA 716
Db 217 AlaThrSerThrArgCysTrpProAlaAlaSerAlaAsnSerValProLysProSer 236
QY 717 ACCTGCATATCG-----TGACCTGTGTCGGGACCCCC 749
Db 237 SerProArgSerValArgAlaGlyThrArgTrpGlyCysCysThrTrp-----ThrPro 254
QY 750 GGGCGCTGCTTCGTTCCCGAGAACGACACAAAGGAGATCTCATGATTGACATGCGCAT 809
Db 255 GlyAlaProAlaAlaProArg----- 261
QY 810 TGATGGGCGAGCATGAGCAAAAACCAAGAGAGGAGGACCAACCTTACTATGTGATGC--- 866
Db 262 -----AlaThrThrArgAsnSer-----CysSer 270
QY 867 AGTCTATCTGCCAACCCAGCTGGAGA----- 893
Db 271 HisSerProSerArgArgValTrpArgLeuThrThrProGlySerThrProSerArgPro 290
QY 894 ---TCTACAGACCATCCAGTCTTGC-----CCAAGCCCTGCGAGGAACGCTACCTGC 944
Db 291 AlaSerAlaArgProProArgArgCysAlaArgProAlaProCysTrpArgAlaAlaCys 310
QY 945 TTGTGCGCTATGAGG 959
Db 311 ThrCysProThrArg 315

RESULT 15
CGR15
collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
C/Species: Rattus norvegicus (Norway rat)
C/Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C/Accession: A90559; A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209; A91
R/Bornstein, P.
Biochemistry 8, 63-71, 1969
A/Title: Comparative sequence studies of rat skin and tendon collagen. II. The absence of
A/Reference number: A90559; MUID:69155173; PMID:5777344
A/Contents: CNBR0 and CNBR1
A/Accession: A90559
A/Molecule type: protein
A/Residues: 1-19 <BO1>
A/Experimental source: tendon

A;Note: sequences from skin and tendon appear to be identical
A;Note: the amino-terminal tetrapeptide may be removed by limited proteolysis during ext
R;Kang, A.H.; Bornstein, P.; Piez, K.A.
Biochemistry 6, 788-795, 1967
A;Title: The amino acid sequence of peptides from the cross-linking region of rat skin c
A;Reference number: A90552; MUID:67162268; PMID:5337886
A;Contents: CNBr1
A;Accession: A90552
A;Molecule type: protein
A;Residues: 5-19 <BA>
A;Experimental source: skin
R;Bornstein, P.
J. Biol. Chem. 242, 2572-2574, 1967
A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.
A;Reference number: A92029; MUID:67165368; PMID:4290711
A;Contents: CNBr2
A;Accession: A92029
A;Molecule type: protein
A;Residues: 20-55 <BO2>
A;Experimental source: skin and tendon
R;Butler, W.T.; Ponds, S.L.
Biochemistry 10, 2076-2081, 1971
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90353; MUID:71263178; PMID:4327399
A;Contents: CNBr4
A;Accession: A90353
A;Molecule type: protein
A;Residues: 56-102 <BU1>
A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cov
A;Reference number: A90566; MUID:70085124; PMID:5411206
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <BU2>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alphas-1, 1971
A;Reference number: A90357; MUID:72136131; PMID:4335087
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <BA1>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Hermanson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alphas-1, 1972
A;Reference number: A90362; MUID:73006942; PMID:4342027
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <BA2>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2946-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90379; MUID:74271384; PMID:436532
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein
A;Residues: 419-567 <BU3>
A;Experimental source: skin
R;Stoltz, M.; Timpi, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neut
A;Reference number: A91209; MUID:74011954; PMID:4126850
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>

A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) in
R;Stoltz, M.; Timpi, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin
A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequen
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (<
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin colla
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F;1/Modified site: blocked amino end (Glx); probably pyrrolidone carboxylic acid) #statu
F;9/Modified site: allysine (Lys) #status experimental
F;103,424,547/Binding site: carboxylate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
Alignment Scores:
Pred. No.: 0.0279 Length: 671
Score: 138.00 Matches: 101
Percent Similarity: 28.50% Conservative: 19
Best Local Similarity: 23.99% Mismatches: 137
Query Match: 3.73% Indels: 165
DB: 1 Gaps: 23
US-09-645-078-1 (1-2043) x CORT-S (1-671)
QY 1144 GGGCAAGACCAGCCAGCCTGGGAGACATTAAGGCATCCCTGGCAT----- 1095
Db 80 GlyGlnArgGlyProGlyProGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 99
QY 1094 -----TTGCTGGAAAGCGGTGCACCCATGC 1068
Db 100 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAsnThrGly 119
QY 1067 CCTTGCTCGGTGATGTTATGCACCAGCTCTGAAGATGGGCA-----AGA 1020
Db 120 ProAla-----GlyProLysGlyGluProGlySerProGlyGluAsnGly 134
QY 1019 ATTCCAAATCCACGAATTCATACATTCGGGAAGTCTGGCCACAGGGCTC----- 969
Db 135 AlaPro-----GlyGlnMetGlyProArgGlyLeuProGlyGlu 147
QY 968 -----GAGCCAGGTCTCT-----CATAGCCCAACAGCA 942
Db 148 ArgGlyArgProGlyProGlySerAlaGlyAlaArgGlyAspGlyAlaValGly 167
QY 941 GTTAGCGTTCTCGAGGCGCTTGGCAAGAGTGAATGCTTTGTAGATCTCCAGCTGGC 882
Db 168 AlaAlaGlyProGlyProGlyProThrGlyProThrGly-----ProProGly 182
QY 881 TTGGCAGATGACCTGCATCATAGTAGGGTGTGCTCT----- 843
Db 183 PheProGlyAlaAlaGlyAlaLysGlyGluAlaGlyProGlnGlyAlaArgGlySerGlu 202
QY 842 ---CCTTCTTGTAGTTTGTCTCATGCTGCCCATCAATGCGACTGTCAATCATGAGAT 786
Db 203 GlyProGlnGlyValArgGlyGluProGlyProGly----- 215
QY 785 CTCCTTTGTGCGTCTTCGGGAACGAAACACGGCCGGGGGTCCCGGA---CCAGTGGCA 729
Db 216 ---ProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAla 234
QY 728 CGATATGCAGGTTGAGGGAGGGGTCTTTCAGCAGCG-----GCTAGAGGG 684

Search completed: May 6, 2004, 11:07:20
Job time : 63.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:52:38 ; Search time 23.5 Seconds
(without alignments)

9053.560 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattcattgtgttggtga.....tgggatccttaaaaaaaa 2043

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Fgapop 6.0			7.0
Delop 6.0			7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/p/US09645078/runat_06052004_104559_21727/app_query.fasta_1.2183
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPEL=0 -IOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pt0 -NORM=ext -HSPSIZE=500 -MINLEN=0 -WAXLEN=200000000
-USER=US09645078 -CGEN 1_22 @runat_06052004_104559_21727 -NCFU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500.5	13.4	458	1	C6ST_CHICK
2	147	4.0	660	1	YHL1_EBV
3	147	4.0	1453	1	C111_MOUSE
4	145	3.9	1464	1	C111_HUMAN
5	143.5	3.8	1172	1	TSP2_HUMAN
6	141	3.8	1460	1	C111_CANFA
7	140.5	3.8	1003	1	MBD6_HUMAN
8	139.5	3.8	1464	1	C133_MOUSE
9	138.5	3.7	1372	1	CA21_MOUSE
10	138.5	3.7	1418	1	CA12_HUMAN
11	138	3.7	671	1	C111_RAT
12	138	3.7	1453	1	C111_CHICK
13	137.5	3.7	1233	1	MUSA_HUMAN
14	137.5	3.7	1496	1	CA25_HUMAN
15	137	3.7	1366	1	CA21_CANFA
16	136	3.6	830	1	SREC_HUMAN
17	136	3.7	1049	1	CA13_BOVIN
18	135	3.6	784	1	YAV2_XANCV

C 19	134	3.6	1964	1	NTC4_MOUSE	P31695 mus musculus
C 20	133.5	3.6	1459	1	CA12_MOUSE	P28481 mus musculus
C 21	133	3.6	825	1	ICP0_HSV2H	P28284 herpes simp
C 22	133	3.6	1372	1	CA21_RAT	P02466 rattus norv
C 23	132.5	3.6	1355	1	CA21_CANCA	O42350 rana catesb
C 24	132.5	3.6	2944	1	CA17_HUMAN	O02388 homo sapien
C 25	131.5	3.6	1364	1	CA21_BOVIN	P02465 bos taurus
C 26	131	3.5	636	1	CA13_RAT	P13941 rattus norv
C 27	131	3.5	2003	1	NTC4_HUMAN	Q99466 homo sapien
C 28	130.5	3.5	1700	1	BAR3_CHITE	Q03376 chironomus
C 29	130.5	3.5	2003	1	NTC4_HUMAN	Q99466 homo sapien
C 30	130	3.5	1175	1	HCN4_RABIT	Q9TV66 oryctolagus
C 31	128	3.5	1362	1	CA21_CHICK	P02467 gallus gall
C 32	127	3.4	830	1	SREC_HUMAN	Q14162 homo sapien
C 33	127	3.4	1516	1	CA1H_HUMAN	P39060 homo sapien
C 34	126	3.4	1712	1	CA24_HUMAN	P08572 homo sapien
C 35	125.5	3.4	210	1	KR47_HUMAN	P08572 homo sapien
C 36	125.5	3.4	316	1	CC07_CAEL	P18832 caenorhabdi
C 37	125.5	3.4	964	1	IF2_COREF	O8FPA7 corynebacte
C 38	125	3.4	680	1	CA1A_HUMAN	O03592 homo sapien
C 39	125	3.4	1366	1	CA21_HUMAN	P08123 homo sapien
C 40	125	3.4	1466	1	CA13_HUMAN	P02461 homo sapien
C 41	124	3.4	1033	1	IF2_STRCO	Q8CJQ8 streptomyce
C 42	124	3.4	2318	1	NTC3_MOUSE	Q61982 mus musculus
C 43	123	3.3	1046	1	IF2_STRAW	O82K53 streptomyce
C 44	122.5	3.3	1804	1	CA1B_MOUSE	Q61245 mus musculus
C 45	122	3.3	627	1	SPD2_NEPL	P46804 nephila cla

ALIGNMENTS

RESULT 1

C6ST_CHICK STANDARD; PRT; 458 AA.

AC Q92179; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).
OS Gallus gallus (Chicken)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.
RC STRAIN=White Leghorn; TISSUE=Embryonic chondrocytes;
RX MEDLINE=95355490; PubMed=7629189; Kato M., Kimata K.,
RA Fukuta M., Uchimura K., Nakashima K.,
RA Shinomura T., Habuchi O.;
RT "Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase."
RL J. Biol. Chem. 270:18575-18580(1995).
CC -1- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN SULFATE.
CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin = adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D49915; BAA08655.1; -
CC PIR; A57397; A57397.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC Transferase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.


```
QY 1211 CGTGGCGGTAGCCAGCAAAATTCATGCGATCGCCACAGCGCTTCTGAAGTCGAGAAACCT 1152
Db 256 ArgSerGlyAlaAlaAlaGlnArgThrHisArg 268
QY 1151 TTTCATAGGCAAGACACAGCGCCAGCCTGGG 1110
Db 269 ProGlyCysProArgSerAlaAlaAsnProGlyCysProArgThrTrpArgArgSer 289
QY 1109 GGGCATCCCTGGCAATTTGTGTGAAGCGTGTGTCACCCATGCCCTTGCCTGGGTGATGT 1050
Db 289 Gly 300
QY 1049 TATGACCCAGGTCTGAGATGGGCGCAAGAAATTCATCCACGAATTCATACATTCGGG 990
Db 301 GlnArgProSerGlyProThrGlyGlyArgProAlaAlaPro 314
QY 989 AAGTCTGGGCCACAGCGGCTCGAGCCAGCTCTCTCATAGCGCACAGCGTAGCGTTCTCT 930
Db 315 GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaValPro 332
QY 929 GGAGGG 876
Db 333 SerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAla 352
QY 875 AGATGACCT 858
Db 353 ArgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArg 372
QY 857 AGTAGGGTGTCTCTCT 816
Db 373 CysProAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArg 392
QY 815 CCATCAATGAGTGTCAATCATGAGATCTCCCTTTGTGGTCTC 768
Db 393 ProProGlyCysPro 411
QY 767 741
Db 412 ArgSerGlyAlaGlnArgGlyHisProProGlyAlaGlyGlnArgProSerGlyPro 431
QY 740 GGACAGGTGCAGTATGACAGTGTGAGGAGGGCTCTTCAGCAGG 690
Db 432 ThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 451
QY 689 AGAGGACTGCGAGTGTGAAGAGCGCACCTCTTCAGCACACGCTGGGTGTAGGAGCGGC 630
Db 452 GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGly 469
QY 629 AGCGCTTCTCCA 606
Db 470 ProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuPro 489
QY 605 585
Db 490 GlnAspLeuAlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAla 509
QY 584 TGCAGTGAG 543
Db 510 AlaAlaGlnArgThrHisArgProProGly 521
QY 542 CACAGCGAGTGCAGACACAGGCGCGCTCTCTCCCACT 489
Db 522 ProArgSerAlaArgAsn 539
QY 488 ACTGTCTCCGGGACAGGTTTCATGTAGGCGATCAAGACGCTCATGTGCGCACAGAAGA 429
Db 540 AlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGly 559
QY 428 CGGCGCGTATCATCCGACAGCCATGTGAGCATCCAGCGGTCTCTCTCTGAGGG 369
Db 560 ArgProAlaAla 567
```

```
QY 368 TCATCCACAGTGCAGCGGGCTCATCAGGTAGAAACATCTGGTGTGCC 315
Db 568 GlyThrProAlaAlaProGlyProGlyGly 582
QY 314 CAAAAGCTGCCCAAGAAAGAGCCAGAGCGCCAGGAGACAGAACACCAAGTGC 255
Db 583 SerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProPro 600
QY 254 TGGCTCGGGCTGTGCTTCATAGCAGGAGCTATGTTGGCTGTACATGTGGAAGA 195
Db 601 AlaAlaArgLeuProProGlu 608
QY 194 ATAGACCCAGATGGCCATCTGGGAACACAGAAACAGCAGGAGCTTCATTTT 135
Db 609 GlnGluProArgLeuProGln 620
QY 134 GTAGCATTTGTGTAAGTGTGAAGACCTGTGCTGGCTTGGCTTCCCTGCTGCTTT 75
Db 621 GlnArgCysProAlaGlyProProProThrArgSerGly 633
QY 74 CTCACTGTAGCGGCCCATGCTTTCTCT 36
Db 634 AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnPro 653
QY 35 TTCTGTGCTCTTCCGTACCCACACAAATGG 6
Db 654 GlyCys 659
RESULT 3
CALL_MOUSE STANDARD; PRT; 1453 AA.
AC P11087; Q60635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COL1A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen."
RL Matrix Biol. 14:593-595(1995).
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
collagen protein."
RL Gene 39:311-312(1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
evidence for a mouse B1 element within the gene."
RL Mol. Cell. Biol. 2:1362-1371(1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157103; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences."
RL DNA 1:59-69(1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE=88124276; PubMed=3340560;
```



```

Db 503 ProGly-----504
QY 206 ACATGTGAAGATAGACCAAGATGCCATCTGGAAACACGAAACAGAGAGCTTCA 147
Db 505 -----ProAlaGlyProLysGlySerProGlyGluAla-----515
QY 146 TTTTITTAGCAGTAGCATTCGCTGAAGTGAAGACCTTGTGGCTTGGCTTCCCTGCT 87
Db 516 -----GlyArgProGlyGluAlaGlyLeuProGly 525
QY 86 GCTTTTGAAGAGCTCACTGCTAGCGCGCCATCGCTTTTCTCTCTACCC 36
Db 526 Ala---LysGlyLeuThrGlySerProGly-----SerProGlyPro 538

RESULT 4
CALL HUMAN
ID CALL HUMAN STANDARD; PRT; 1464 AA.
AC P02452; Q14037; Q15176;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT chain of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [2]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [3]
RP SEQUENCE OF 162-301.
RX TISSUE=Skin;
RC MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [4]
RP SEQUENCE OF 263-268.
RX TISSUE=Skin;
RC MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT "A comparative study of glycopeptides derived from selected
RT vertebrate collagens. A possible role of the carbohydrate in fibril
RT formation.";
RL J. Biol. Chem. 245:5042-5048(1970).
RN [5]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;
RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN [6]
RP SEQUENCE OF 1229-1454 FROM N.A.
RX TISSUE=Bone;

```

```

RX MEDLINE=88124208; PubMed=3340531;
RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RL domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to
RT transcriptional control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RL cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalgleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN [15]
RP VARIANT OI-II ARG-569.
RX MEDLINE=87222295; PubMed=3109247;
RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RL collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
RN [16]
RP VARIANT OI-II CYS-926.
RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;

```



```

FT DISULFID 558 574 BY SIMILARITY.
FT DISULFID 577 588 BY SIMILARITY.
FT DISULFID 594 610 BY SIMILARITY.
FT DISULFID 601 619 BY SIMILARITY.
FT DISULFID 622 646 BY SIMILARITY.
FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 707 715 BY SIMILARITY.
FT DISULFID 720 740 BY SIMILARITY.
FT DISULFID 756 776 BY SIMILARITY.
FT DISULFID 779 799 BY SIMILARITY.
FT DISULFID 815 835 BY SIMILARITY.
FT DISULFID 838 858 BY SIMILARITY.
FT DISULFID 876 896 BY SIMILARITY.
FT DISULFID 912 932 BY SIMILARITY.
FT DISULFID 948 1169 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;

Alignment Scores:
Pred. No.: 0.014 Length: 1172
Score: 143.50 Matches: 131
Percent Similarity: 29.09% Conservative: 61
Best Local Similarity: 19.85% Mismatches: 229
Query Match: 3.83% Indels: 239
DB: 1 Gaps: 31

US-09-645-078-1 (1-2043) x TSP2_HUMAN (1-1172)
QY 199 CCACATGACAGCCACCAATCAGCTCCCTGTCTATGAAGGCAGCCGCGAG----- 250
Db 248 ProHisValThrThrGluThrValGlyPro-SerSerGluArgArgProGluValCysgl 267
QY 251 -----CGCATGCCAGCTCGT 267
Db 267 uArgSerCysGluGluLeuGlyAsnMetValGlnGluLeuSerGlyLeuHisValLeuVa 287
QY 268 T-----CTGCTCTCTCGGCGCTCTGGCTCTCTTTTGTGGGCGAGCT 309
Db 287 lAsnGlnLeuSerGluAsnLeuLysArgValSerAsnAspAsnGlnPheLeuTrpGluLe 307
QY 310 TTTTGGGCGAGCCAGATGTTTCTACCTGATGGAGCCCGCTCGCAGCTGTGGATGAC 369
Db 307 uileGlyGlyProProLysThr---ArgAsnMetSerAlaCysTrpGlnAspGlyArgph 326
QY 370 CTTCAAGCAGACAGCCGCTGGATGCTGCACATGGCTGTGCGGATCTGATACGGGCGGT 429
Db 326 ePheAlaGluAsnGluThrTrp-----ValValAspSerCy 338
QY 430 CTTCTTGTGGCAGATGAGCGCTTTTGATGCTCCTACATGNAACCTGCTCCCGGAGACATC 489
Db 338 sThrThrCysThrCysLysLysPhe----- 346
QY 490 CAGCCTCTTTCAGTGGGAGACAGCGCGCCCTGTGT-----TCTGCACCTGCGCTG 540
Db 347 -----LysThrileCysHisGlnIleThrCysPro-ProA 358
QY 541 TGACATCATCCCAAGATGAATATATCCCGGGCTCACTCCAGCTCTCTGT----- 593
Db 358 lathrCys-----AlaSerProSerPheValGluGluCysCysPros 373
QY 594 -----GCAGTCAACAGCCCTTTCAGTGTGGTGGAGAGGCTGCGCTCTCTACAG 642
Db 373 erCysLeuHisSerValAspGlyGluGluGlyTrp-----SerProTrpA 388
QY 643 CCACGTGG-----TGCTCAAGAGAGGTGCGCTTCTTCAACCTGCAGTCCCTCTACCCGCT 696

```

```

Db 388 laGluTrpThrGlnCysSerValThrCysGlySerGlyThrGlnGlnArgGlyArgSerC 408
QY 697 GTTGAAGAGCCCTCCCTCAACCTGCATA-----TCGTGCACC- 734
Db 408 yaAspValThr---SerAsnThrCysLeuGlyProSerIleGlnThrArgAlaCysSerL 427
QY 735 -----TGCT 738
Db 427 euSerLysCysAspThrArgIleArgGlnAspGlyGlyTrpSerHisTrpSerProTrps 447
QY 739 CCGGGACCCCGCGGTTCCTCCGAGACGCAAGAGGAGATCTCATGATTGA 798
Db 447 erSerCysSerValThrCysGlyValGlyAsnIleThrArgIleArgLeuCysAsnSerP 467
QY 799 CAGTCGCATTGTGATGGGCGCAGCATGAGCAAAATCAAGAGAGGAGGACCAACCTACTA 858
Db 467 roValProGlnMetGlyGlyLysAsnCysLysGlySerGlyArgGluThrLysAlaCysG 487
QY 859 TGTGATGCGAGTCACTGCGCAAGCCAGCTGAGATCTACAAGACCATCTCTGCGC 918
Db 487 lGlyAlaProCysProIleAspGlyArgTrpSerProTrpSerProTrpSerAlaCysT 507
QY 919 CRAAGGCCCTGC-----AGGACGCTACCTGTGTGCGCTATGAGGACCTGCGC 966
Db 507 hrValThrCysAlaGlyGlyIleArgGluArgThrArgValCys----- 521
QY 967 TCAGAGCCCTGGCCCGACACTTCCGAA-----TGTATGAATT 1005
Db 522 -----AsnSerProGluProGlnTrpGlyGlyLysAlaCysValGlyA 536
QY 1006 CG-----TGGGATTGGAATTCTTTCGCCCATCTTCAGACCTGGGTGCA 1047
Db 536 spValGlnGluArgGlnMetCysAsnLysArgSerCysProVal-----AspGlyCysL 554
QY 1048 TAACATCACCCGAGGCAAGGCGATGGGTGACCACGCTTCCACACAATGCCAGGATGC 1107
Db 554 euSerAsnPro-----CysPheProGlyAlaG 563
QY 1108 CCTTAATGCTCCAGGCTTGGCGCT-----GGTCTTTGCCCTATGAAAAGGTTTC 1158
Db 563 lncCysSerSerPheProAspGlySerTrpSerCysGlyPheCysProVal---GlyPheL 582
QY 1159 TC-----GACTTCAGAAAGCTGTGGCGATGCGCATGAATTGCT 1197
Db 582 euGlyAsnGlyThrHisCysGluAspLeuAspGluCysAlaLeuValProAspIleCysP 602
QY 1198 GGCTACCCGCGCAGCTCAGATCTGAACAAGACAGAGAAACCTGTGCTGGATCTTCTGTC 1257
Db 602 he-----SerThrSerLysValProArgCysValAsnThrGlnProGlyPheHisCysL 620
QY 1258 TACCT-----GGACTGTCCTGAGCA 1278
Db 620 euProCysProProArgTrpArgGlyAsnGlnProValGlyValGlyLeuGluAlaAla 640
QY 1279 AATCCACTAAGAGGTTGAGAAGGCTTGTGCTGCCACCTGGTGCAGCTCAGCTCACTTTC 1338
Db 640 ysthrGluLysGlnValCysGluProGluAsnPro----- 651
QY 1339 TCTGAATGCTTCTGAGCCTTGCTCATCTGAGCCTTAACATACATGCTCTGCGGTATC 1398
Db 652 -----CysLysAspLysThrHisAsnCysHisLysH 662
QY 1399 AACTGAGTGTGAGTTGTGTCCACAGCTGCTCAAGCAGAGAGGACTTTTGTGCTCATGCTT 1458
Db 662 lAlaGluCysIleLeuGlyHisPhe-SerAspProMet-TyrLysCysGluCys-- 680
QY 1459 GTGCTCTAGAAAACAGACTGGGAAACCTTATGTGAGCAGCACATCCACCAGTGAACACAG 1518
Db 681 -----GlnThrGl 683
QY 1519 GTATTGCTCTTCTTCTTCTT-----GATCTTCTGCTGGCAGACTT 1563

```


Db 683 yYrAlaGlyAspGlyLeuIleCysGlyGlyuAspSerAspLeuAspGlyTrpProAsnLe 703
 QY 1564 CAGAGACTTTGGCGCTGGAGGCTATTAAACACAGACACAGATATCACTGGAAATGATCCA 1623
 Db 703 uAsnLeuValCysAlaThrAsnAlaThrTyHis- 716
 QY 1624 TAAACCTCCCTGTCCACATCTTGGCCATGGGAATGGATCTTTCCACAAAGAGCTCAC 1683
 Db 716 eLysAspAsnCysProHisLeuProAsnSerGlyGlnGluAspPheAspLysAspGlyI1 735
 QY 1694 AGCATTTTCCACAGAG- 1699
 Db 736 eGlyAspAlaCysAspAspAspAspAspAspGlyValThrAspGluLysAspAsnCy 756
 QY 1700 -ATGCGAATCTGAGCCCTGAGCTTCCCAATGGATTCAAGGAGGAAGTGGGAACAAG 1758
 Db 756 sGlnLeuLeuPheAsnProAsgGlnAlaAspTyAsp---LysAspGluValGlyAspAr 775
 QY 1759 GTTGATGCCCTACTATGAGCTTGACCATAGCTATCGGTAATCAGAAATA 1810
 Db 775 gCysAspAsnCys- 788
 RESULT 6
 CAIL CANFA
 ID CAIL CANFA STANDARD; PRT; 1460 AA.
 AC Q9XSJ7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1
 OS Canis familiaris (Dog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 CC NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
 RC TISSUE=Skin;
 RX MEDLINE=21023337; PubMed=11147834;
 RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
 RT "Sequence of normal canine COL1A1 cDNA and identification of a
 RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case
 RT of canine osteogenesis imperfecta";
 RL Arch. Biochem. Biophys. 384:37-46(2000).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 CC (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimer of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
 CC (OI).
 CC -!- SIMILARITY: Contains 1 WFCC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF153062; AAD34619.1; -
 CC InterPro; IPR008161; C1g helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR000885; Fib collagen_C.
 CC InterPro; IPR002181; Fibrinogen_C.
 CC InterPro; IPR001007; WFC_C.
 CC Pfam; PF01410; COLFI; 1.
 CC Pfam; PF01391; Collagen; 18.
 CC ProDom; PD000007; C1g helix; 2.
 CC ProDom; PD002078; Fib collagen_C; 1.
 CC SMART; SM00038; COLFI; 1.
 CC SMART; SM00214; WVC; 1.

DR PROSITE; PS01208; WFCC_1; 1.
 DR PROSITE; PS0184; WFCC_2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Disease mutation.
 FT SIGNAL 1 22
 FT PROPEP 23 157
 FT CHAIN 158 1214
 FT PROPEP 1215 1460
 FT DOMAIN 34 92
 FT DOMAIN 158 174
 FT DOMAIN 175 1198
 FT DOMAIN 1169 1214
 FT SITE 741 743
 FT SITE 1089 1091
 FT CARBOHYD 1361 1361
 FT VARIANT 208 208
 SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;
 Alignment Scores:
 Pred. No.: 0.0186 Length: 1460
 Score: 141.00 Matches: 102
 Percent Similarity: 30.68% Conservative: 25
 Best Local Similarity: 24.64% Mismatches: 137
 Query Match: 3.81% Indels: 151
 DB: 1 Gaps: 24
 US-09-645-078-1 (1-2043) x CAIL CANFA (1-1460)
 QY 1144 GGGCAAGACAGCGCCAAAGCCCTGGAGACATTAAGGGCATCCCTGGCAT-----TTG 1092
 Db 238 GlyGluArgGlyProGlyProGlyGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 257
 QY 1091 TGTGGAAACGCTGCTAC----- 1074
 Db 258 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGly 277
 QY 1073 CCATGCCCTTCCCTCGGCTGATGTTATGCACCCAGCTCTGAAGATGGGCGCAAGAATTCCA 1014
 Db 278 ProAlaGlyProLysGlyGluProGlySerPro---GlyGluAsnGlyAlaPro----- 294
 QY 1013 ATCCCACGAATTCATATCGGGAAGTCGGGCCACAGGGCTC----- 969
 Db 295 -----GlyGlnMetGlyProArgGlyLeuProGlyGluArgGly 307
 QY 968 -----GAGCCAGGTCTCTCATACGCGCACAA-----GCAGGTAGCCTTCTCTGCA 927
 Db 308 ArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAla 327
 QY 926 GGGCCTTGGCAAGAGCTGGATGGTCTTGTAGATCTCCAGCTGGCTTGGCAGATGACCT 867
 Db 328 GlyProGlyProThrGlyProAla-----GlyProGlyPheProGlyAlaVal 345
 QY 866 GCATCATAGTAGGGTTGGTCT-----CCTTCTTGAGTT 831
 Db 346 GlyAlaLysGlyGluAlaGlyProGlnGlyAlaArgGlySerGluGlyProGlnGlyVal 365
 QY 830 TTTGCTCATGCTCCCTCCATCACAATCGACTGTCAATCATGATCTCCCTTTGCGGTT 771
 Db 366 ArgGlyGluProGlyProGly-----ProAlaGlyAla 377
 QY 770 CTGGGNAAGGAACACGGCCCGGGTCCCGGA-----CCAGGTGCAGATATGAGGTTGA 714
 Db 378 AlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAlaLysGlyAlaAsnGly 397
 QY 713 GGGAGGGGTCTTTCCAGCAGCG-----GGTAGAGGGACTCGAGTTGAAGA 669
 Db 398 AlaProGlyIleAlaGlyAlaProGlyPheProGlyAlaArgGlyProSerGlyProGln 417
 QY 668 -----AGCCACCTCTTCCAGCACACACGTGGCTGTAGAGCGCGAGGCTTCTCCACCA 615
 Db 418 GlyProSerGlyProPro-----GlyProLysGly---AsnSerGlyGluProGlyAlaPro 435
 QY 614 CCTCAAGGGCTGTTGACTGCACAGAGGCTGCAGTGAGCCCGGGGATGATTCATCTT 555

```

Db 436 GlyAsnLysGly-----AspThrGLYAlaLysGlyGluProGly----- 448
QY 554 GTGGGATGATGTACAGGAGGTGAGAAACAGAGGCGCGGCTGTCTCCCACTGAAGA 495
Db 449 -----ProThr----- 450
QY 494 GCGTGAGTCTCTCGGGACAGGTTCCATGTAGGATCAA---AGACGTCATGTCGC 438
Db 451 -----GlyIleGlnGlyProProGlyProAlaGlyGluGlyLysArgGlyAlaArg 468
QY 437 ACAAGAGACGCGCGTATCAGATCCCGCACAGACATGTGCAGCATCCAGCGGTGCTCT 378
Db 469 GlyGluProGlyProThrGlyLeuProGlyProPro----- 480
QY 377 GCTTGAAGTTCATCCACAGTCCAGCGCGGCTCCATCAGTAGAACAACATCTGGTGTCT 318
Db 481 -----GlyGluArgGlyGlyProGlySerArgGlyPheProGlyAla 494
QY 317 GCCCAAAAGTGTCCCAAAAGAGAGAGCCAGAGCGCCAGAGACAGACAGACGACGT 258
Db 495 AspGlyValAlaGlyPro-----LysGlyProAlaGlyGluArgGlySerProGly--- 511
QY 257 GCATCGCTCGGCTGTGCTTTCATAGACAGGAGCTGATGTTGTGCTGTACATGTGA 198
Db 511 ----- 511
QY 197 AGAATAGAGCAAGATGCCATCTCGGAAACCCAGAAACAGAGGAGCTTCATTTTTTAG 138
Db 512 -----ProAlaGlyProLysGlySerProGlyGluAla----- 522
QY 137 GCAGTAGCATGTGTGAAGTGAAGACCTGTGTGGCTTGGCTCCCTGCTGCTTTTGAG 78
Db 523 -----GlyArgProGlyGluAlaGlyLeuProGlyAla---Lys 534
QY 77 AGGCTCACTGTAGCGCGGCGCATGCGCTTTCTCTCTTACCC 36
Db 535 GlyLeuThrGlySerProGly-----SerProGlyPro 545

RESULT 7
MBD6 HUMAN
ID MBD6 HUMAN STANDARD; PRT; 1003 AA.
AC Q96N6; Q8N3M0; Q8N81; Q96Q00;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Methyl-CpG binding domain protein 6.
GN MBD6 OR KIAA1887.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagaetsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
```

```

RL DNA Res. 8:179-187 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Amygdala;
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK056399; BAB71176.1; -
DR EMBL; AK093078; BAC04045.1; -
DR EMBL; AB067474; BAB67780.1; -
DR EMBL; AL834230; CAD38908.1; -
DR Genew; HGNC:20445; MBD6.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
FT DOMAIN 17 61 MBD.
FT DOMAIN 140 846 PRO-RICH.
FT CONFLICT 637 637 D -> G (IN REF. 1; BAB71176).
FT CONFLICT 802 802 Q -> QQ (IN REF. 1; BAC04045).
SQ SEQUENCE 1003 AA; 101200 MW; 6C8B8693AA6A3BE6 CRC64;

Alignment Scores:
Pred. No.: 0.081 Length: 1003
Score: 140.50 Matches: 137
Percent Similarity: 32.11% Conservative: 38
Best Local Similarity: 25.14% Mismatches: 156
Query Match: 3.75% Indels: 216
DB: 1 Gaps: 30

US-09-645-078-1 (1-2043) x MBD6_HUMAN (1-1003)
QY 64 CTACAGTGTAGCTCTCAAAAGCAGGAGGAG-----CCNAGCCACAGGT 111
Db 319 LeuAlaSerSerLeuLeuSerAlaAlaLysAlaGlnHisProLeuProPro 338
QY 112 CTTCCACTTCAGCACAACTACTGCTCTAAATAAATGAAGTCTCTTCTGTTTC 171
Db 339 SerThrLeuGlnGlyArgPro-----ArgAlaGlnAlaProSerAlaSer 354
QY 172 CCAGATGGCCATCTTGGCTCTATTCTTCACATGTACAGCCACACATGCTCCCT 228
Db 355 -----HisSerSerLeuArgProSerGlnArgProArgProProThr 371
QY 229 -----GTCTATGAGGCACAGCC-----CGAGCCGATGACGCTGCT 264
Db 372 ValPheArgLeuLeuGlyArgGlyProGlnThrProArgSerArgProArgAla 391
QY 265 GGTCTCTGT-----CTTCTGCG 281
Db 392 ProAlaProValProGlnProPheSerLeuProGluProSerGlnProLeuProSer 411
QY 282 GCTCTGCTCTCTCT-----TTGTGGGCGGACGCTTT 311
Db 412 ValLeuSerLeuGlyLeuProThrProGlyProSerHisSerAspGlySerPheAsn 431
QY 312 ---TTGGGCGAGCACCCAGATGTTTCTACCTGATGAGCCCGCTCGACGTGTGGATGA 368
Db 432 LeuLeuGlySerAspAlaHisLeu-----ProProProProThrLeu----- 445
QY 369 CTTTCAAGCAGACACCGCTGTGCTGACATGCTGTGCGGATCTGATACGGCGCG 428
Db 446 ---SerSerGlySerProPro----- 451
QY 429 TCTTCTTGTGCGACATGAGCGTCTTTGATGCTTACATGGAACCTGGTCCCGGAGACGT 488
```

Db 452 -----GlnProArgHisProIleGlnProSerProGlyLeuProGlyThr 465
QY CCAGCCTTTTCACT -----GGGAGACACCGGGCCCTGTCTGCACCTGCCT 539
Db 466 SerGlySerLeuSerSerValProGlyAlaProAlaProAlaSerLysAlaPro 485
QY GTGACATCATCCCAAGATGAATCATCCCCC -----GGGCTCACTGCAGGCTCC 590
Db 486 ValValProSerProValLeuGlnSerProSerGluGlyLeuGlyMetGlyAlaGlyPro 505
QY 591 TGTGACGTCAACAGCCCTTGTGAGTGTTGGAGAGCCCTGCGCTCTACAGCACCTGG 650
Db 506 AlaCysProLeuProProLeuAlaGlyGlyGluAlaPheProPheProSerProGlu-- 524
QY 651 TGCTCAAGAGGTGGCTT -----CTTCAAGCTGCAGTCCCTCT 689
Db 525 ---GlnGlyLeuAlaLeuSerGlyAlaGlyPheProGlyMetLeuGlyAlaLeuProLe 543
QY 690 ACCC ----- 693
Db 543 uProLeuSerLeuGlyGlnProProProSerProLeuLeuAsnHisSerLeuPheGlyVa 563
QY 694 -----GCTGTGAAGACCCCTCCCTCAACCTGCATATCGTCACCT---GGTCC 740
Db 563 lLeuThrGlyGlyGlyGlnProProProGlnProLeuLeuProProProGlyGlyPr 583
QY 741 GGGACCCCGGGCGCTGTCTCCGTCGCCGAGAACGACAAAGGAGATCTCATGATTGACA 800
Db 583 oGlyProPro ----- 586
QY 801 GTGCGATTGTGATGGCGCAGCATGAGCAAAACTCAAGAGGAGCAACCCCTACTATG 860
Db 587 -----LeuAlaProGlyGluProGluGlyProSerLeuLeuVa 599
QY 861 TGATCCAGTCACTGCCAAGCCAGCTGGAGATCTCAAGACCAT ----- 906
Db 599 l---AlaSerLeuLeuProProProProSerAspLeuLeuProProSerAlaProPr 618
QY 907 -----CCAGTCTCTGCCAAGCCCT---CGAGAACGCT 938
Db 618 sSerAsnLeuLeuAlaSerPheLeuProLeuLeuAlaLeuGlyProThrAlaGlyAsp-- 637
QY 939 ACCTGCTTTGTGGCTATGAGGACCTGGCTCGAGC -----CCCTGTGCCCC 983
Db 638 -----GlyGluGlySerAlaGluGlyAlaGlyGlyProSerGlyG1 651
QY 984 AGACTTCCGGAATGATGAATCGTGGATGGATTTCTTCCCA ----- 1029
Db 651 uProPhe -----SerGlyLeuGlyAspLeuSerProLeuLeuPheProPr 666
QY 1030 -----TCTTCAGACCTGGGTGCATTAACATCA 1055
Db 666 cLeuSerAlaProProThrLeuLeuAlaLeuAsnSerAlaLeuLeuAlaThrLeuAs 686
QY 1086 CCGAGGCAAGGCGATGGTGACACCGCTTCCACAAATGCCAGGATGCCCTTAATG 1115
Db 686 pProSerGlyThrPro -----ProGlnProCys ----- 696
QY 1116 TCTCCAGGCTTGGCGCTGGTCTTCCCTATGAAAGGTTTCTCGACTTCA ----- 1167
Db 697 -----ValLeuSerAlaProGlnProGlyProProThrSerSerValTh 711
QY 1168 -----GAAAGCTGTGGCGATGCCCATGAATTTGCTGGGCT 1202
Db 711 rThrAlaThrAspProGlyAlaSerSerLeuGlyLysAlaProSerAsnSerGlyAr 731
QY 1203 ACCGCCAGCTCAGATCTCAAGAACAGACAGAACTCTGTGATCTTCTGCTACCT 1262
Db 731 gProProGlnLeuLeu -----SerProLeuLeuGlyAlaSerLeuLeuG1 746
QY 1263 GGAAGTGTGAGCAATCACTAAGAGGCTTGAAGAGGCTTGTGCTGCCACCTGG 1318

746 yAspLeuSerSerLeuThrSerSerProGlyAlaLeuProSerLeuLeuGlnProProG1 766
1319 TGTTCAGCTCACT 1331
Db 766 yProLeuLeuSer 770
RESULT 8
CA13_MOUSE
ID CA13_MOUSE STANDARD; PRT; 1464 AA.
AC P08121; Q61429; O9CRN7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
RA Tomian D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
complete DNA sequence.";
RL Gene 147:161-168(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Richards S., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Bjchalons D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Hulyk S.W.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogeli G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
alpha-1 type-III collagen chain.";
RL Gene 61:225-230(1987).
RN [4]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Liao G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
(III) collagen gene.";
RL J. Biol. Chem. 260:3773-3777(1985).
RN [5]
RP SEQUENCE OF 810-1464 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gusninch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RA "Functional annotation of a full-length mouse cDNA collection";
 RT Nature 409:685-690(2001).
 RN [6]
 RN SEQUENCE OF 1442-1464 FROM N.A.
 RC STRAIN=C57BL/6;
 RC MEDLINE=91274355; PubMed=2054384;
 RA Metasartanta M., Toman D., de Crombrughe B., Vuorio E.,
 RT "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs";
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
 CC along with type I collagen.
 CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
 CC linked to each other by interchain disulfide bonds. Trimers are
 CC also cross-linked via hydroxylsines.
 CC -!- PTM: Proline residues at the third position of the tripeptide
 CC repeating unit (G-X-Y) are hydroxylated in some or all of the
 CC chains.
 CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
 CC the oxygen atom of a post-translationally added hydroxyl group (By
 CC similarity). Contains 1 VWFC domain.
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X52046; CAA36279.1; -;
 DR EMBL; BC043089; AAH43089.1; -;
 DR EMBL; BC058724; AAH58724.1; -;
 DR EMBL; M18933; AAA37338.1; -;
 DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
 DR EMBL; AK019448; BAB31724.1; -;
 DR EMBL; X57983; CAA41048.1; -;
 DR FIR; A27353; A27353.
 DR PIR; S59856; S59856.
 DR MGD; MGI:88453; Col3a1.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib. collagen_C.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR001007; VWFC.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Clg helix; 1.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWFC; 1.
 DR PROSITE; PS01208; VWFC 1; 1.
 DR PROSITE; PS0184; VWFC 2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.

FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 90 VWFC.
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
 FT DOMAIN 1196 1464 NONHELICAL REGION (C-TERMINAL).
 FT CARBOHYD 262 O-LINKED (GAL. .) (BY SIMILARITY).
 FT MOD_RES 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 976 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1093 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1105 HYDROXYLATION (BY SIMILARITY).
 FT DISULFID 1195 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 1196 INTERCHAIN (BY SIMILARITY).
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886030B CRC64;
 Alignment Scores:
 Pred. No.: 0.024 Length: 1464
 Score: 139.50 Matches: 101
 Percent Similarity: 32.87% Conservative: 16
 Best Local Similarity: 28.37% Mismatches: 121
 Query Match: 3.77% Indels: 118
 DB: 1 Gaps: 23
 US-09-645-078-1 (1-2043) x CAL3_MOUSE (1-1464)
 QY 1124 CTTGGAGACATTAAAGGCATCCCTGGCATTTGTGTGGAACGGTGGTCAACCCATGCCCT 1065
 DB 814 ProGlyGlnAsn---GlyGluPro---GlyAlaLysGlyGluArgGlyAlaPro----- 829
 QY 1064 TGCTCGGTGATGTTATGCACCCAGGTCTGAAGATGTGGGCAAGAAATTCATCCACGA 1005
 DB 830 -----GlyGluLysGlyGlyGlyPro-----ProGlyProAla 841
 QY 1004 ATTCATACATTCGGGAAGTCTGGGCACACAGAGGCGTCGAGCCAGGTCCTCAT-----AGC 951
 DB 842 GlyProThr---GlySerSerGlyProAlaGly---ProProGlyProGlnGlyValLys 859
 QY 950 GCACAAAGAGGTAGCTTCTCGAGGCGCTTCGGCAAGGACTGGATGGTCTTGTAGATCT 891
 DB 860 GlyGluArgGlySer---ProGlyGlyPro-----Gly 869
 QY 890 CCAGCTGGCTTT-----GGCAGA-----TGCACCTGCATCATAGTAGGTT 849
 DB 870 ThrAlaGlyPheProGlyGlyArgGlyLeuProGlyProGlyAsnAsnGlyAsnPro 889
 QY 848 GGTCTCTCTTCTAGTCTTTTCTCATGCTGCCCATCAATGCGACTGTCAATCATGA 789
 DB 890 GlyProProGly-----ProSerGlyAlaProGlyLys----- 900
 QY 788 GATCTCCCTTTGTGCTTCTCGGGAAACGGAACACGGCCCGGGGGTCCCGGA----- 738
 DB 901 AspGlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaGly 920
 QY 737 -----CCAGGTGCACGATATGCAGGT 717
 DB 921 ProLysGlyAspAlaGlyGlnProGlyGlyLysGlyProGlyAlaGlnGlyProPro 940
 QY 716 TGAGGAGAGGGT---CTTTCAGCAGCGGGT-----AGCGGACTCGAGTTGA 672
 DB 941 GlySerProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGly--- 959
 QY 671 AGAAGCGCACCTCTTGTAGCACCCAGCTGCTGTAGGAGCGGCGAGGCGCTTCTCCACCACCT 612
 DB 960 -----ProProGlyMetProGlyProArgGlySerProGlyPro 972
 QY 611 CAAAGGCGCTTGTAGCTGCACAGAGCGCTGCAGTGCAGCCCGGGGGATGATTTCTCTGTG 552
 DB 973 Gln-----GlyIleLysGlyGlySerGlyLysProGlyAlaSerGlyHis----- 967
 QY 551 GGATGATGTCACAGGAGGTGCAGAACACAGAGGCGCGGCTGTCTCCCTCAAGAGAGGC 492
 DB 988 -----Asn-GlyGluArgGlyProProGlyProGlnGlyLeuProGlyGln----- 1002

QY 1127 AAGCTGGGACACATTAAAGGCATCCCTGGCATTTGTGTGAAAGCGTGTCTACCCATGC 1068
Db LysProGlyGluArg---GlyHisPro---GlyLeuAlaGlyAlaArgGlyAlaProGly 529
QY 1067 CCTTGCCTCGGTCATGTTATGCACCCAGGTCCTGAAGATGGGCAAGAAATTCCAATCCA 1008
Db Pro----- 530
QY 1007 CGAATTCATACATTCGGGAAGTCTGGGCCACAGGGCTCGAGCCAGGTCTCTATAGCGCA 948
Db -----AspGlyAsnAsnGlyAlaGlnGly---ProProGlyProGlnGlyVal 545
QY 947 CAAGCAGGTAGCTTCCTCGAGGCGCTTGGGCACAGGACTGGATGCTTCTAGATCTCCA 888
Db GlnGlyGlyLysGlyGlnGlnGlyPro-----AlaGlyPro 557
QY 887 GCTGCTTTGGCAGATGACCT-----GCATCACATAGTAGGTTGGTCTCTCTCT 837
Db ProGlyPheGlnGlyLeuProGlyProSerGlyThrThrGlyGluValGlyLysProGly 577
QY 836 TGAGTTTGTCTCATGTCGCCCATCATGCGACTGTCATCATGAGATCTCCCTTTG 777
Db -----GluArgGlyLeuPro----- 582
QY 776 TCGCTTCGCGAAGCGAAGACCGGCGGCGGCTCCCGACAGGTCGACGATATGCAGGT 717
Db -----GlyGluPheGlyLeuProGlyProAlaGlyPro----- 593
QY 716 TGAGGAGGCTTCTTCAGCAGCGGTAGAGGACTGAGGTGAAGA-----AGCGCACT 660
Db -----ArgGly-----GluArgGlyThrProGlyGluSerGlyAlaAlaGly 607
QY 659 CTTGAGCACACAGTGGCTGTAGG-----AGCGCAGGCTTCTCCACCACTCAAGG 606
Db ProSerGlyProIleGlySerArgGlyProSerGlyAlaProGlyProAspGlyAsnLys 627
QY 605 GCTGTTGACTGCACAGGACCTGCAGTGCAGCCGCGGAGATTCATCTTGTGGGATGA 546
Db GlyGlu-----AlaGlyAlaValGlyAlaProGly----- 637
QY 545 TGTCACAGGAGTCTCAGACACAGGCGCGCTTCTCCACTGAAAGAGGTGGACT 486
Db -----SerAlaGlyAlaSerGlyProGlyGlyLeuProGlyGluArgGlyAlaAl 654
QY 485 G-----TCTCCGGGACCAAGTTCATGTAGGCATCAAA-----GACGCTCATGT 441
Db acGlyileProGlyGlyLysGlyGluLysGlyGluThrGlyLeuArgGlyAspThrGlyAs 674
QY 440 CGCAAGACAGCGCGGTATCATCCGCACAGCCATGTCAGCATCCAGCGCGGTGC 381
Db nThrGlyArgAspGlyAlaArgGlyileProGlyAla---ValGlyAlaProGlyProAl 693
QY 380 TCTGCTTGAAGTTCAT-----CCACAGTGCAGCGCGGCT 345
Db aglyAlaSerGlyAspArgGlyGluAlaGlyAlaAlaGlyProSerGlyProAlaGlyPr 713
QY 344 -----CATCAGGTAGAAACATCTGGT----- 321
Db oArgGlySerProGlyGluArgGlyGluValGlyProAlaGlyProAsnGlyPheAlaGl 733
QY 320 -----GTCGCCCAAAAGCTGCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 274
Db yProAlaGlyAlaAlaGlyGlnProGlyAlaLysGlyGlyLysGlyThrLysGlyProLy 753
QY 273 GACAGAACAGCAGCTGTCATCGCTCGGCTGTGCTTCATAGACAGGAGCTGATGTTG 214
Db sglyGlu-----AsnGlyileValGlyProThrGlySer----- 764
QY 213 TGGCTGTACATGTGGAAGATAGAGCCAGATGGCCATCTGGGAACACAGCAAGCAGG 154
Db -----ValGlyAlaAlaGlyProSerGlyProAsnGlyProProGlyProValGl 781
QY 153 AGCTTCATTTTATGAGCTAGCATGTGCTGAAAGTGGAACCTTGTGGCTTG---GG 97

Db 781 yserArgGly----- 793
QY 96 CTTCCCTCTCTCTTTTGAGAGG 75
Db 793 yPheProGlyAlaAlaGlyArg 800
RESULT 10
CAL2 HUMAN
ID CAL2 HUMAN STANDARD; PRT; 1418 AA.
AC P02458;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
GN COL2A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90067946; PubMed=2587267;
RA Su W.W., Lee B., Ramirez F., Machado M., Horton W.;
RT "Nucleotide sequence of the full length cDNA encoding for human type II procollagen.";
RL Nucleic Acids Res. 17:9473-9473(1989).
RN [2]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=87031574; PubMed=3021582;
RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
RT "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL Gene 44:11-16(1986).
RN [3]
RP SEQUENCE OF 432-1145 FROM N.A.
RA Ramirez F.;
RL Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 963-1418 FROM N.A.
RX MEDLINE=85190534; PubMed=3857598;
RA Cheah K.S.E., Stoker N.G., Griffin J.R., Grosfeld F.G., Solomon E.;
RT "Identification and characterization of the human type II collagen gene (COL2A1).";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN [5]
RP SEQUENCE OF 1120-1398 FROM N.A.
RX MEDLINE=85306861; PubMed=3840017;
RA Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,
Vuorio E.;
RT "Construction and identification of a cDNA clone for human type II procollagen mRNA.";
RL Biochem. J. 229:183-188(1985).
RN [6]
RP SEQUENCE OF 1106-1418 FROM N.A.
RX MEDLINE=88067771; PubMed=2825137;
RA Elima K., Vuorio T., Vuorio E.;
RT "Determination of the single polyadenylation site of the human pro alpha 1(II) collagen gene.";
RL Nucleic Acids Res. 15:9499-9504(1987).
RN [7]
RP SEQUENCE OF 1227-1289 FROM N.A.
RX MEDLINE=86104139; PubMed=3002437;
RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for a human pro-alpha 1(II) collagen chain and demonstration of restriction fragment length polymorphism at the 3' end of the gene.";
RL Biochemistry 24:6343-6348(1985).
RN [8]
RP SEQUENCE OF 1176-1226 FROM N.A.
RX MEDLINE=84118798; PubMed=6320112;
RA Strom C.M., Upholt W.B.;
RT "Isolation and characterization of genomic clones corresponding to the human type II procollagen gene.";

RL Nucleic Acids Res. 12:1025-1038(1984).
 RN [9]
 RP SEQUENCE OF 35-167 FROM N.A.
 RX MEDLINE=89233138; PubMed=2714801;
 RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
 RT "Organization of the exons coding for pro alpha 1(II) collagen N-
 RT propeptide confirms a distinct evolutionary history of this domain of
 RT the fibrillar collagen genes";
 RL Genomics 4:438-441(1989).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans";
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97259599; PubMed=9101280;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP VARIANT SER-1074.
 RX MEDLINE=90036309; PubMed=2572591;
 RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
 RA Hollister D.W.;
 RT "Glycine to serine substitution in the triple helical domain of pro-
 RT alpha 1 (II) collagen results in a lethal perinatal form of short-
 RT limbed dwarfism";
 RL J. Biol. Chem. 264:18265-18267(1989).
 RN [13]
 RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.
 RX MEDLINE=89266907; PubMed=2543071;
 RA Lee B., Vissing H., Ramirez F., Rogers D., Rimooin D.;
 RT "Identification of the molecular defect in a family with
 RT spondyloepiphyseal dysplasia";
 RL Science 244:978-980(1989).
 RN [14]
 RP VARIANT OSTEOARTHRITIS CVS-650.
 RX MEDLINE=90370826; PubMed=1975693;
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a
 RT cause of primary osteoarthritis associated with a mild
 RT chondrodysplasia";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
 RN [15]
 RP VARIANT OI-IV VAL-717.
 RX MEDLINE=91291136; PubMed=2064612;
 RA Bateman J.F., Hamman M., Chan D., Cole W.G.;
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
 RT valine substitution in osteogenesis imperfecta type IV. Detection of
 RT the mutation and prenatal diagnosis by a chemical cleavage method.";
 RL Biochem. J. 276:765-770(1991).
 RN [16]
 RP VARIANT OSTEOARTHRITIS CVS-650.
 RX MEDLINE=91086471; PubMed=1985108;
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;
 RT "Cartilage expression of a type II collagen mutation in an inherited
 RT form of osteoarthritis associated with a mild chondrodysplasia";
 RL J. Clin. Invest. 87:357-361(1991).
 RN [17]
 RP VARIANT HYPOCHONDROGENESIS GIU-984.
 RX MEDLINE=93054548; PubMed=1429602;
 RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimooin D.L.,
 RA Cohn D.H., Eyre D.R.;
 RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
 RT 1(II) chain produces hypochondrogenesis";
 RL J. Biol. Chem. 267:22522-22526(1992).
 RN [18]
 RP VARIANT HYPOCHONDROGENESIS SER-705.

RX MEDLINE=92262484; PubMed=1374906;
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
 RA Ramirez F., Vitale E., Lee B.;
 RT "Characterization of a type II collagen gene (COL2A1) mutation
 RT identified in cultured chondrocytes from human hypochondrogenesis";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
 RN [19]
 RP VARIANT WS-II ASP-198.
 RX MEDLINE=93304428; PubMed=8317498;
 RA Koerkoe J., Ritvanemi P., Haataja L., Kaeeriaeinen H.,
 RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
 RT for glycine alpha 1-67 and that causes cataracts and retinal
 RT detachment: evidence for molecular heterogeneity in the Wagner
 RT syndrome and the Stickler syndrome (arthro-ophthalmopathy)";
 RL Am. J. Hum. Genet. 53:55-61(1993).
 RN [20]
 RP VARIANT SEMD CYS-840.
 RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimooin D.L.,
 RA Eyre D.R.;
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces
 RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
 RL Am. J. Hum. Genet. 53:A209-A209(1993).
 RN [21]
 RP VARIANT OSTEOARTHRITIS CVS-650.
 RX MEDLINE=93282819; PubMed=8507190;
 RA Holderbaum D., Mailemud C.J., Moskowitz R.W., Haqqi T.M.;
 RT "Human cartilage from late stage familial osteoarthritis transcribes
 RT type II collagen mRNA encoding a cysteine in position 519";
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
 RN [22]
 RP VARIANT SEMD ARG-285.
 RX MEDLINE=93252400; PubMed=8486375;
 RA Vikkula M., Ritvanemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
 RA Peltonen L.;
 RT "A mutation in the amino-terminal end of the triple helix of type II
 RT collagen causing severe osteochondrodysplasia";
 RL Genomics 16:282-285(1993).
 RN [23]
 RP VARIANT SEDC CYS-206.
 RX MEDLINE=94063862; PubMed=8244341;
 RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
 RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
 RT family with an Arg75-->Cys mutation in the procollagen type II gene
 RT (COL2A1)";
 RL Hum. Genet. 92:499-505(1993).
 RN [24]
 RP VARIANT SEDC CYS-920.
 RX MEDLINE=93315508; PubMed=8325895;
 RA Chan D., Taylor T.K.F., Cole W.G.;
 RT "Characterization of an arginine 789 to cysteine substitution in
 RT alpha 1(II) collagen chains of a patient with spondyloepiphyseal
 RT dysplasia";
 RL J. Biol. Chem. 268:15238-15245(1993).
 RN [25]
 RP VARIANT SEDC SER-1128.
 RX MEDLINE=93140139; PubMed=8423604;
 RA Cole W.G., Hall R.K., Rogers J.G.;
 RT "The clinical features of spondyloepiphyseal dysplasia congenita
 RT resulting from the substitution of glycine 997 by serine in the alpha
 RT 1(II) chain of type II collagen";
 RL J. Med. Genet. 30:27-35(1993).

Alignment Scores:

Pred. No.:	0.0282	Length:	1418
Scores:	138.50	Matches:	87
Percent Similarity:	35.47%	Conservative:	18
Best Local Similarity:	29.39%	Mismatches:	119
Query Match:	3.74%	Indels:	72
DB:	1	Gaps:	17

US-09-645-078-1 (1-2043) x CA12_HUMAN (1-1418)


```
Oy 1142 GCAAGACCGCGCC---AAGCCTGGGAGACATTAAAGGCATCCCTGGGCATTTGTGTGA 1086
Db      |||||
Db 349 AlalysGlySerAlaGlyAlaProGly-----ileAlaGly 360
Oy 1085 AAGCGTGTCACCCATCCCTGCTGCTGGTGATGTTATGCACCCAGGTCTGAAGATGG 1026
Db      |||||
Db 361 AlaProGlyPheProGlyPro-----ArgGlyProProAspProGlnGly 375
Oy 1025 GCAGAATTCATC---
Db      |||||
Db 376 AlaThrGlyProLeuGlyProLysGlyGlnThrGlyLysProGlyileAlaGlyPheLys 395
Oy 992 GGGAGCTGGCCACAGGCTCAGCAGGTCTCATAGCGCACAGCAGGTAGCGTT 933
Db      |||||
Db 396 GlyGlnGlnGlyProLysGly---GluProGlyProAlaGlyProGlnGlyAlaProGly 414
Oy 932 CCTGACGGCTCTGGCAAGGACTGCTGTTGTAGATCTCGAGTGGCTTTGGCAGA 873
Db      |||||
Db 415 ProAlaGlyGluGluGlyLysArgGly---AlaArgGlyGluProGlyGlyValGly--- 432
Oy 872 TGACCTGCATACATAGTAGGTGGTCTCTCTCTTGGTTTGGT----- 825
Db      |||||
Db 433 -----ProileGlyProProGlyGluArgGlyAlaProGlyAsnArg 446
Oy 824 -----CATGCTGCCCATCACAATGCGACTGT 798
Db      |||||
Db 447 GlyPheProGlyGlnAspGlyLeuAlaGlyProLysGlyAlaProGlyGluArgGlyPro 456
Oy 797 CAATCATGAGATCTCCCTTTGCTCGGTAACGGAACACCGCGGGGGTCCCGGA 738
Db      |||||
Db 467 SerGlyLeuAlaGlyProLysGlyAlaAsnGlyAspProGlyArgProGlyGluProGly 486
Oy 737 ---CCAGGTGCACGATATGCAGGTGAGGAGGGCTCTTTCCAGCAGCGGTAGAGGACT 681
Db      |||||
Db 487 LeuProGlyAlaArgGlyLeuThr---GlyArgProGlyAspAlaGlyProGlnGlyLys 505
Oy 680 CGAGTTGAAGAGCGACCTCTCTTGACACACACGCTGCTAGGACGGCGGCTTCT 621
Db      |||||
Db 506 ValGlyProSerGlyAlaProGly-----GluAspGlyArgProGly 519
Oy 620 CCACCA---CCTCAAGGGCTGTTGACTGCACAGGAGCTGACGTGAGCCGGGGATGA 564
Db      |||||
Db 520 ProProGlyProGln-----GlyAlaArgGlyGlnProGly----- 531
Oy 563 TTTCATCTGTGGGATGATGTCACAGGAGGTGCAGAACACAGGCGCGGCTGTTCTCC 504
Db      |||||
Db 532 -----ValMetGlyPheProGlyProLysGlyAlaAsnGlyGluProGlyLysAlaGly 549
Oy 503 ACTGAAGAGCGTGGACTGCTCCGGGACAGGTTCCATGTCAGCATCAAGACGCTCA 444
Db      |||||
Db 550 Glu-LysGlyLeuProGlyAlaProGlyLeuArgGlyLeuProGly---LysAspGlyGly 569
Oy 443 TGTGCAACAAGAGCGCGGTATCAGATC-----CGCACAGCCCATGTGCA 396
Db      |||||
Db 568 uThrGlyAlaGluGlyProProGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGly 588
Oy 395 GCATCAGCGCGTGTCTGTCGTGAAGTCATCCACAGTGCAGGC 350
Db      |||||
Db 589 yAlaProGlyProSerGlyPheGlnGlyLeuProGlyProProGly 603

RESULT 11.
ID _CALL11 RAT STANDARD; PRT; 671 AA.
AC P02454; P02455;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain (Fragments).
GN COL1A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

```
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-19.
RX MEDLINE=69155173; PubMed=577344;
RA Bornstein P.;
RT "Comparative sequence studies of rat skin and tendon collagen. II.
RT The absence of a short sequence at the amino terminus of the skin
RT alpha-1 chain.";
RL Biochemistry 8:63-71(1969).
RN [2]
RP SEQUENCE OF 5-19.
RX MEDLINE=67162268; PubMed=5337886;
RA Kang A.H., Bornstein P., Piez K.A.;
RT "The amino acid sequence of peptides from the cross-linking region of
RT rat skin collagen.";
RL Biochemistry 6:788-795(1967).
RN [3]
RP SEQUENCE OF 20-55.
RX MEDLINE=67165368; PubMed=4290711;
RA Bornstein P.;
RT "The incomplete hydroxylation of individual prolyl residues in
RT collagen.";
RL J. Biol. Chem. 242:2572-2574(1967).
RN [4]
RP SEQUENCE OF 56-102.
RX MEDLINE=71263178; PubMed=4327399;
RA Butler W.T., Ponds S.L.;
RT "Chemical studies on the cyanogen bromide peptides of rat skin
RT collagen. Amino acid sequence of alpha 1-CB4.";
RL Biochemistry 10:2076-2081(1971).
RN [5]
RP SEQUENCE OF 103-139.
RX MEDLINE=70085124; PubMed=5411206;
RA Butler W.T.;
RT "Chemical studies on the cyanogen bromide peptides of rat skin
RT collagen. The covalent structure of alpha 1-CB5, the major
RT hexose-containing cyanogen bromide peptide of alpha 1.";
RL Biochemistry 9:44-50(1970).
RN [6]
RP SEQUENCE OF 140-238.
RX MEDLINE=72136131; PubMed=4335087;
RA Balian G., Click E.M., Bornstein P.;
RT "Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of
RT the hydroxylamine-produced fragment HA1.";
RL Biochemistry 10:4470-4478(1971).
RN [7]
RP SEQUENCE OF 239-418.
RX MEDLINE=73006942; PubMed=4342027;
RA Balian G., Click E.M., Hermodson M.A., Bornstein P.;
RT "Structure of rat skin collagen alpha 1-CBB. Amino acid sequence of
RT the hydroxyl amine-produced fragment HA2.";
RL Biochemistry 11:3798-3806(1972).
RN [8]
RP SEQUENCE OF 419-567.
RX MEDLINE=74271984; PubMed=4366532;
RA Butler W.T., Underwood S.P., Finch J.E. Jr.;
RT "Chemical studies on the cyanogen bromide peptides of rat skin
RT collagen. Amino acid sequence of alpha 1-CB3.";
RL Biochemistry 13:2946-2953(1974).
RN [9]
RP SEQUENCE OF 568-651.
RX MEDLINE=74011954; PubMed=4326850;
RA Stoltz M., Timpi R., Furthmayr H., Kuehn K.;
RT "Structural and immunogenic properties of a major antigenic
RT determinant in neutral salt-extracted rat-skin collagen.";
RL Eur. J. Biochem. 37:287-294(1973).
RN [10]
RP SEQUENCE OF 651-671.
RX MEDLINE=73049495; PubMed=4636751;
RA Stoltz M., Timpi R., Kuehn K.;
RT "Non-helical regions in rat collagen alpha 1-chain.";
RL FEBS Lett. 26:61-65(1972).
RN [11]
```



```

QY 38 CCC 36
DB 387 PRO 387

RESULT 12
CALL CHICK
ID CALL CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RN [1]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=88056316; PubMed=3678834;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RT of the chicken pro alpha 1(I) collagen mRNA.";
RL Gene 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA Kang A.H., Gross J.;
RT "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP SEQUENCE OF 152-1187.
RX MEDLINE=82231995; PubMed=7093229;
RA Highberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Kang A.H., Gross J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glimcher M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP SEQUENCE OF 981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP SEQUENCE OF 1311-1453 FROM N.A.
RX MEDLINE=80134546; PubMed=6987088;
RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA Pastan J., Decrombrugne B., Fietzek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT carboxyl end of pro alpha 1(I)-chains.";
RL FEBS Lett. 111:61-65(1980).
RN [7]
RP FUNCTION: Type I collagen is a member of group I collagen
RP (fibrillar forming collagen).
CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -1- PTM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the

```

```

CC chains. Pro-1153 is the only 3-hydroxypro and the only
CC hydroxylated proline in position X.
CC -1- SIMILARITY: Contains 1 VWFC domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M17839; AAA48704.1; -.
CC EMBL; M17838; AAA48704.1; JOINED.
CC EMBL; V00401; CAA23695.1; -.
CC EMBL; M10571; AAA48671.1; ALT_SEQ.
CC EMBL; M17607; AAA48672.1; -.
CC PIR; A21179; A27179.
CC PIR; I50629; I50629.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC InterPro; IPR002181; Fibrinogen_C.
CC InterPro; IPR001007; VWFC_C.
CC Pfam; PF01410; COLFI_1.
CC Pfam; PF01331; Collagen; 18.
CC Pfam; PF00033; vwc; 1.
CC ProDom; PD000007; Clg_helix; 2.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC SMART; SM00214; VWC; 1.
CC PROSITE; PS01208; VWFC_1; 1.
CC PROSITE; PS50184; VWFC_2; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
CC SIGNAL 1 22
CC PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
CC CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.
CC PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
CC DOMAIN 31 89 VWFC
CC MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
CC MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
CC MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
CC MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
CC MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
CC MOD_RES 1153 1153 HYDROXYLATION.
CC CONFLICT 1187 1187 F -> L (IN REF. 5).
CC CONFLICT 1441 1441 Q -> H (IN REF. 6).
CC SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Alignment Scores:
Pred. No.: 0.0309 Length: 1453
Score: 138.00 Matches: 112
Percent Similarity: 31.31% Conservative: 22
Best Local Similarity: 26.17% Mismatches: 156
Query Match: 3.73% Indels: 129
DB: 1 Gaps: 25

US-09-645-078-1 (1-2043) x CALL_CHICK (1-1453)
QY 1144 GGGCAAGACAGCGCCAGCCTGGGAGACATTAAGGCGATCCTCGCAT-----TTG 1092
DB 114 GlyAspArgGlyLeuProGlyProProGlyArgAspGlyLeuProGlyGlnProGlyLeu 133
QY 1091 TGTGGAAGCGTGTGTCACCA---TGCCCTGCTCGGGTGATGTTATGACCCAGG--- 1038
DB 134 ProGlyProProGlyProProGlyProProGlyLeuGlyGlyAsnPheAlaProGlnMet 153
QY 1037 -----TCTGAAGATGGGCAAGATTCCTCCATCCAGCAATTCATACATTCGGG 990
DB 154 SerTyrGlyTyrAspGluLysSerAlaGlyValAlaValPro-----GlyProMetGly 171
QY 989 AAGTCTGGGCCACAGGGGGCTC-----GAGCCAGGTCTCTCATGCGCACAA 945

```


DR GO; GO:0005201; Extracellular matrix structural constituent; TAS.
 DR GO; GO:0007155; P:cell adhesion; NAS.
 DR InterPro; IPR006208; Cys_knot.
 DR InterPro; IPR006207; Cys_knot_C.
 DR InterPro; IPR002919; TIL_Cysrich.
 DR InterPro; IPR001007; VWF_C.
 DR InterPro; IPR001846; VWF_D.
 DR Pfam; PF00007; Cys_knot; 1.
 DR Pfam; PF00093; vwc; 1.
 DR Pfam; PF00094; vwd; 1.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PROSITE; PS01208; VWF_1; 2.
 DR PROSITE; PS0184; VWF_2; 2.
 KW Repeat; Glycoprotein.
 FT NON_TER 1
 FT DOMAIN 431 496 VWF 1.
 FT DOMAIN 500 659 VWF.
 FT DOMAIN 688 712 5 X 5 AA TANDEM REPEATS OF T-T-V-G-P/S.
 FT REPEAT 688 692 1.
 FT REPEAT 692 697 2.
 FT REPEAT 697 702 3.
 FT REPEAT 702 707 4.
 FT REPEAT 707 712 5.
 FT DOMAIN 857 923 VWF 2.
 FT DOMAIN 962 1026 VWF 3.
 FT DOMAIN 1111 1199 CTCK.
 FT DISULFID 1111 1161 BY SIMILARITY.
 FT DISULFID 1136 1175 BY SIMILARITY.
 FT DISULFID 1140 1191 BY SIMILARITY.
 FT DISULFID 1158 1193 BY SIMILARITY.
 FT DISULFID 1160 1198 BY SIMILARITY.
 FT CONFLICT 193 194 VS -> HE (IN REF. 2).
 FT CONFLICT 406 406 P -> R (IN REF. 4).
 FT CONFLICT 463 465 SER -> REP (IN REF. 2).
 FT CONFLICT 478 478 G -> A (IN REF. 2).
 FT CONFLICT 492 594 HHYQCVCVSGWGDHYITFDQTYTFLDNCTYVLVQGI
 FT VYGHVLDVNFVCGAEDGLSCPSIIILEXHODRVLLTRK
 FT VHGVTNIIENKVVSEGR -> ITTSASVCAAGVTPT
 FT TSPSTAPTPSTTARTIGADCARVPLPRARRQLLRG
 FT GRALLPEVHHPGVPPGADPQASPRGVGDKRDLHQGGQ
 FT PLPL (IN REF. 2).
 FT G -> A (IN REF. 4).
 FT H -> Q (IN REF. 3).
 FT A -> R (IN REF. 2).
 FT GIMCFPTPADRVYQFCG -> RTCMHLPSRQGVPLR
 FT (IN REF. 2).
 FT R -> P (IN REF. 4).
 FT S -> T (IN REF. 3).
 FT G -> R (IN REF. 2).
 FT A -> R (IN REF. 2).
 FT G -> R (IN REF. 2).
 FT SQCCGTCVQACVINTSKSPAHLPYGE -> RRSVAPY
 FT CRSPVSPPARAPPTSTLAS (IN REF. 2).
 FT L -> P (IN REF. 2).
 FT C -> S (IN REF. 2).
 FT GTOHSEAEPEPSQEAESGWGVPVSPMH -> ATPEST
 FT RRRPSPARRORVAGREASSVPHALS (IN REF. 2).
 FT VSPMH -> QCPPTDCHRPDLDQGFPIPISSASKAS
 FT GTCAPVQAAALNTLSFAFLWRVWANGHLIPGGALTHPA
 FT CSHLSGAPGLAEILLWPCIQPAVLGT (IN REF. 4).
 FT SEQUENCE 1233 AA; 130072 MW; 37CEA5FADCB8A28DE CRC64;
 Alignment Scores:
 Pred. No.: 0.032 Length: 1233
 Score: 137.50 Matches: 123
 Percent Similarity: 32.76% Conservative: 48
 Best Local Similarity: 23.56% Mismatches: 197
 Query Match: 3.67% Indels: 154
 DB: 1 Gaps: 24

US-09-645-078-1 (1-2043) x MUSA_HUMAN (1-1233)

QY 426 CCGTCTTCTGTGGACATGAGCGCTTTGATGCTACATGAACCTG-----GTCCCC 479
 D 29 ProSerGlyProGlyThrThrProSerProValProThrThrSerThrThrSerAlaPro 48
 QY 480 GGAGACAGTCCAGCGCTCTTTCAGTGGGAGAAACAGCCGGCCCTGTGTCTGCACCTGCT 539
 D 49 ThrThrSerThrThrSerGlyProGlyThrThrProSerPro-----ValPro 64
 QY 540 GTGACATCA-----TCCCAACAAGATGAATCATCTCCCGGGGCTC 578
 D 65 ThrThrSerThrThrSerAlaProThrThrThrThrThrSerAlaSerThrAlaSerThr 84
 QY 579 ACTGCAAGGCTCTGTGTCAGTCAACAGCCCTTTAGGTGGTGGAGAGAGCGCTGCCCTCT 638
 D 85 ThrSerGlyProGlyThrThrProSerProValProThrThrThrThrThrSerAlaPro 104
 QY 639 ACAGCCACGCTGTGCTCAAGAGGTGCGCTTCTCACTGAGTCCCTCTACCCGCTGC 698
 D 105 ThrThrThrThrThrSerAlaSerThrAlaSerThrSerGlyProGlyThrSerLeu 124
 QY 699 TGAAGAGCCCT---CCCTCAACCTGCATATCGTGCACCTGTCGGGACCCCGGGCG 755
 D 125 SerProValProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 144
 QY 756 TGTTCCTTCCCGAGAACCCAAAGGAGATCTCATGATTGACAGTCCGATGTGATGG 815
 D 145 GlyThrThrProSerProVal----- 151
 QY 816 GGCAAGTCCAGCAAACTCAAGAGGAGGACCACTACTATGATGAGGTCATCT 875
 D 152 -----ProThrThrThrThrThrAla 159
 QY 876 GCCAAGCCAGCTGGAGATCTCAAGACATCCATGCTTCCCAAGGCGCTGCAAGAAC 935
 D 160 ProThrThrThrThrSerGlyProGlyThrThrProSerProValProThrThrSer 179
 QY 936 GCTACC----- 941
 D 180 ThrThrProValSerLysThrSerThrHisLeuSerValSerLysThrThrHisSer 199
 QY 942 -----TGC-----TTGTGCGTATGAGGACCTGGCTGCGAGCCCTG 977
 D 200 GlnProValThrSerAspCysHisProLeuCysAlaTrpThrLysTrpPheAspValAsp 219
 QY 978 TGCCCCAGACTT---CCGAATGTATGAATTCGTGGATGGAATCT----- 1022
 D 220 PheProSerProGlyProHisGlyAspLysGluThrTyrrAsnAsnIleIleArgSer 239
 QY 1023 -----TGCCCATCTTCAGACCTGCGTGCATCAACATCACCCGAGCAAGGCA 1070
 D 240 GlyGluLysIleCysArg-----ArgProGluGluIleThrArgLeuGlnCysArgAla 257
 QY 1071 TGGGTG---ACCAGCTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTT 1127
 D 258 GluSerHisProGluValAsnIleGluHisLeuGlyGlnValGlnCysSerArgGlu 277
 QY 1128 GCGCTGCTTTC-----CCTATGAAAAGGTTTCTCGACTTC 1166
 D 278 GluGlyLeuValCysArgAsnGlnAspGlnGlnGlyProPheLysMetCysLeuAsnTy 297
 QY 1167 AGAAGCGCTGCGGATGCCATGAATTCGTGGCTACCCGACGTCAGATCTGAACAAG 1226
 D 298 GluValArgVal-----LeuCysGluThr-ProArgGlyCys----- 310
 QY 1227 AACAGAGAACCTTGTGCTGGATCTTCTGTCTACCTGGATCTCCCTGAGCAATCCACT 1286
 D 311 -----ProValThrSerValThrProTyrrGlyThrSerPro---ThrAsnAlaLe 326
 QY 1287 AAGAGGGTTGAGAGGCTTGTGTCACCTGCTGTCAGCTCAGTCACTTCTCTGAATG 1346
 D 326 u-----TyrProSerLeuSerThrSerMetValSerAlaSerVal----- 339


```

DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib. collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 19.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; Clg_helix; 5.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS0184; VWC_2; 1.
DR ExPASy; PS0184; VWC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
KW Disease mutation; 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 1226
FT PROPEP 1227 1496
FT DOMAIN 39 97
FT MOD_RES 230 290
FT MOD_RES 293 293
FT MOD_RES 296 296
FT MOD_RES 608 608
FT MOD_RES 614 614
FT VARIANT 960 960
FT CONFLICT 292 292
FT CONFLICT 1418 1418
FT CONFLICT 1438 1438
FT CONFLICT 1460 1460
FT CONFLICT 1496 1496
FT SEQUENCE 1496 AA; 144720 MW; 82827C17A86644F5A CRC64;

Alignment Scores:
Pred. No.: 0.034 Length: 1496
Score: 137.50 Matches: 88
Percent Similarity: 33.44% Conservative: 12
Best Local Similarity: 29.43% Mismatches: 107
Query Match: 3.72% Indels: 92
DB: 1 Gaps: 21

US-09-645-078-1 (1-2043) x CA25_HUMAN (1-1496)
QY 1136 ACCAGCGCAAGCTTGGGAGACATTAAAGGCATCTCGTCATTGTGGAAAGCGTGGT 1077
DB 439 ThrSerGlyProGlySer-----AlaGlyProGly 450
QY 1076 CACCCATGCCCT-----TGCTCGGTGATGTTATGCACCGAGTCTGAAGATGGGCA 1023
DB 451 SerProGlyProGlySerThrGly-----ProGlyAsnSerGly--- 465
QY 1022 AGAATTCGAATCCACGAATTCATCATTC---GGAGTCTGGCCACAGGGGCTCGAG 966
DB 466 -----LeuProGlyAspProGlyPheLysGlyGluAlaGlyProLysGly---Glu 481
QY 965 CCAGGCTCTCATAGCCACAGCAGGATAGCGTCTCTGAGGCGCTTGGCAGAGGACTGGA 906
DB 482 ProGlyProHisGlyIleGlyGly-----ProIleGlyProGlyGluGly 498
QY 905 TGCTCTGTAGATCTCCAGTGGCTTGGCAGATGACCTGCATCATAGTAGTGGTGGT 846
DB 499 -----LysArgGlyProArgGly-----AspProGlyThrLeuGly 510
QY 845 CTCTCTTCTGAGTTTTGTCTATGCTGCGCCATCAATGCGACTGTCAATCATGAGAT 786
DB 511 ProProGlyProValGlyGluArgGlyAlaProGlyAsnArgGlyPheProGlySerAsp 530
QY 785 -----CTCCCT 780
DB 531 GlyLeuProGlyProLysGlyAlaGlyGluArgGlyProValGlySerSerGlyPro 550
QY 779 TTGTGCTTCTCGGACCGAAGCAACACCGCGCGGGTCCCGGA---CCAGTGTGACCATAT 723

```

```

Db 551 LysGlySerGlnGlyAspProGlyArgProGlyGluProGlyLeuProGlyAla----- 568
QY 722 GCAGGTTGAGGAGGGGTCTTTTCAGCA-----GCGGGTAGAGGAGTGC 678
Db 569 -----ArgGlyLeuThrGlyAsnProGlyValGlnGlyProGlyLysLeu 584
QY 677 GTTGAAGAAGCGCACCTCTTTCAGCACCACCGTGGCTGTAGAGCGGCGGCTCTTCCCA 618
Db 585 GlyProLeuGlyAlaProGly-----GluAspGlyArgProGlyPro 598
QY 617 CCACCTCAAGAGGCTGTGACTGCACAGGAGCGCTGCAGTGCAGCGGGGATGATTTTCAT 558
Db 599 Pro-----GlySerIleGlyLeuGlyGlnProGly----- 609
QY 557 CTTGTGGGATGATGCACAGGAGGTGCAGAACACAGCGCGCGCTGTCTCCACTGAA 498
Db 610 ThrMetGlyLeuProGlyProLysGlySerAsnGlyAspProGly---LysProGlyGlu 628
QY 497 AGAGCTGGACT---GTCTCGGGGACCGAGTTCATGATGAGCATCAAGAGCGTCAATGT 441
Db 629 AlaGlyAsnProGlyVal-ProGlyGlnArgGlyAlaProGly---LysAspGlyLysVa 647
QY 440 CGCACAGAAGACGCGCGTATCATGATCCCG-----CACAG 405
Db 647 I-----GlyProTyrGlyProGlyProGlyProGlyLeuArgGlyGluArgG 663
QY 404 CATGTGACGATCCAGCGCGTCTCTGTTGAAGTCTCATCCACAGTGCACAGGC 350
Db 663 yGluGlnGlyProGlyProGlyProThrGlyPheGlnGlyHisProGlyProGly 681

RESULT 15
CA21_CANFA
ID CA21_CANFA STANDARD; PRT; 1366 AA.
AC 046352; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF035120; AAC64485.1; .
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 19.
CC ProDom; PD000007; Clg_helix; 3.

```


DR ProDom; PD002078; Fib.collagen_C; 1.
DR SMART; SN00038; COLFI1.1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Collagen; Signal; Pyrrolidone carboxylic acid.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE (BY
FT FT SIMILARITY).
FT CHAIN 80 1102 COLLAGEN ALPHA 2(I) CHAIN
FT PROPEP 1103 1366 CARBOXYL-TERMINAL PROPEPTIDE (BY
FT FT SIMILARITY).
FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID (BY
FT FT SIMILARITY).
FT MOD_RES 84 84 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT FT INVOLVED IN CROSS-LINKING (BY
FT FT SIMILARITY).
SQ SEQUENCE 1366 AA; 129400 MW; CD936969E080BCD4 CRC64;

Alignment Scores:
Pred. No.: 0.036 Length: 1366
Score: 137.00 Matches: 113
Percent Similarity: 30.14% Conservative: 35
Best Local Similarity: 23.01% Mismatches: 191
Query Match: 3.70% Indels: 152
DB: 1 Gaps: 20

US-09-645-078-1 (1-2043) x CA21_CANFA (1-1366)

QY 1180 GCACAGGCTTTCTGAAGTCGAGAAACCTTTTCATAGGCAAGACCGGCAAGCCTG 1121
Db 335 AlaAlaGlyAlaThrGlyAlaArgGlyIleValGlyGluProGlyProAlaGlySerLys' 354
QY 1120 GAGACATT-----AGGGCATCCCTGGCATTTGTGTGAAGCGTGTCCACCTGCC 1067
Db 355 GlyGlySerGlyAsnLysGlyGluProGly--SerAlaGlyAlaGlnGlyProPro----- 372
QY 1066 CTTCCTCGGTCATGTTATGCCACCGAGGTCCTCAAGATGGGCAAGAAATTCCAATCCAC 1007
Db 373 -----GlyProSerGlyGluGlnGlyLysArgGlyProAsnGlyG 386
QY 1006 GAATTCATATCGGGAAGTCTGGCCACAGGGCTCGAGCCAGGTCCTCATAGC----- 951
Db 386 luAlaGlySerAlaGlyProSerGlyProProGlyLeuArgGlySerProGlySerArgG 406
QY 950 -----GCACAGCAGGTAGCCTT---CCTGCAGGCGCTTGGCAAGGA 911
Db 406 lyLeuProGlyAlaAspGlyProAlaGlyValMetGlyProGlyProGlyProAlaGly 426
QY 910 CTGGA---TGCTCTTGTAGATCTCCAGCTGGCTTTGGCAGATCACCTGCATCATAGTA 854
Db 426 hrGlyProAlaGlyValArgGlyProAsnGlyAspSerGlyArgProGlyGluProGlyL 446
QY 853 GGGTTGCTCTCTCTTGTAGTTTGTCTCATGCTGCCCATCACAAATGGAGTGTCAAT 794
Db 446 euMetGlyProArgGly-----PheProGlyAlaProGlyAsnValGlyProAlaG 463
QY 793 CATGAGATCTCCCTTTGTGGTCTCGGGAACGGAACACGCGCCCGGGGT----- 744
Db 463 lyLysGluGlyProMetGlyLeuProGlyIleAspGlyArgProGlyProIleGlyProA 483
QY 743 -----CCCGACCGAGGTGCACGATATGCAG 719
Db 483 laGlyAlaArgGlyGluProGlyAsnIleGlyPheProGlyProLysGlyProThrGlyA 503
QY 718 GTTGAGGGAGGGCTTTTCAGCAGCGGTAGAGGACTGCAGGTTGAAGAAGCGCACCT- 660
Db 503 spProGlyLysAsnGlyAspLysGlyHisAlaGlyLeuAlaGlyAlaArgGlyAlaProG 523
QY 659 -----CCTTGAGCACCAAGCTGGCTGTAGGAGCGGCA 629
Db 523 lyProAspGlyAsnAspGlyAlaGlnGlyProGlyProGlnGlyValGlnGlyL 543
QY 628 GG-----CCTTCTCCACCACT-----CAAAGGCTGTTGACT 596
Db 628 GG-----CCTTCTCCACCACT-----CAAAGGCTGTTGACT 596

Search completed: May 6, 2004, 11:02:17
Job time : 55.5 secs

Db 543 ysGlyGluGlnGlyProAlaGlyProProGlyPheGlnGlyLeuProGlyProAlaGlyT 563
QY 595 GCACAGAGGCTGCAGTGAAGCCCGGG-----GGATGATTTTCATCTGTGG 551
Db 563 hrAlaGlyGluValGlyProGlyGluArgGlyLeuProGlyGluPheGlyLeuProG 583
QY 550 GATGATGTCAGGCGAGTGCAGAAC----- 525
Db 583 lyProAlaGlyProArgGlyGluArgGlyProProGlyGluSerGlyAlaAlaGlyProS 603
QY 525 ----- 525
Db 603 erGlyProIleGlySerArgGlyProSerGlyProProGlyProAspGlyAsnLysGlyG 623
QY 524 -----ACAGGCGCGCGTGTCTTCTCCCA 503
Db 623 luProGlyValLeuGlyAlaProGlyThrAlaGlyAlaSerGlyProGlyGlyLeuProG 643
QY 502 CTGAAGAGGCTGAGCTGTCTCGGGAGCACGAGTTCATGATGATGATCAAGAGCTCAT 443
Db 643 lyGluArgGly-----AlaAlaGly-----IleProGlyGlyLysGlyGluLysGlyG 659
QY 442 GTCCACAGAGAGCGCCGTATCAGATCCCGCAG-----CCAT 401
Db 659 luThrGlyLeuArgGlyGluIleGlyAsnProGlyArgAspGlyAlaArgGlyAlaProG 679
QY 400 GTGCAGCATCCAGCGGTGCTCTGTTGAAGTCCATCCACAGTGCAGCGCGGCTCCAT 341
Db 679 lyAlaMetGlyAlaProGlyProAlaGlyAlaThrGlyAspArgGlyGluAlaGlyProA 699
QY 340 CAGCT-----AGAAAACATCTGGGTG 320
Db 699 laGlyProAlaGlyProAlaGlyProArgGlyThrProGlyGluArgGlyGluValGlyP 719
QY 319 CTGCC-----CAAAAAGCTGCCCAAAAGAGAGAGCCAGA 284
Db 719 roAlaGlyProAsnGlyPheAlaGlyProAlaGlyAlaAlaGlyGlnProGlyAlaLysG 739
QY 283 CGCCAGAGAGACAGAACCCAGCAGTCGCTCGGCTGCTGCTTCATAGCAGGGA 224
Db 739 lyGluArgGlyThrLysGlyProLysGlyGluAsnGlyProValGlyPro---ThrGly- 757
QY 223 GCTGATGTTGTGGTGTACATGTGAAAGATAGAGCAAGATGGCCATCTGGGAAACCAG 164
Db 758 -----ProIleGlySerAlaGlyProSerGlyProAsnGlyProProG 772
QY 163 AAACAGCAGGAGCTTCATTTTTTTAGCAGTAGCATTTGCTGAAGTGGAGAGACCTTGTG 104
Db 772 lyProAlaGlySerArgGlyAsp-----GlyGlyProProGlyA 785
QY 103 GCTTGGGCTTCCCTGCTGCTTTTGAGAGG 75
Db 785 laThrGlyPheProGlyAlaAlaGlyArg 794

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:55:53 ; Search time 104 Seconds
(without alignments)
12396.225 Million cell updates/sec.

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattccattgttgggta.....tgggattcctaaaaa 2043

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cpn2.1/USPTO_spool_p/US09645078/runat_06052004_104559_21739/app_query.fasta_1.2183
-DB=SPTRMBL_25 -QWTF=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09645078 @CNC 1 1 112 @runat_06052004_104559_21739 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	54.5	386	4 Q9Y5R3	Q9Y5R3 homo sapien

ALIGNMENTS	
RESULT 1	
Q9Y5R3	
ID Q9Y5R3	PRELIMINARY; PRT; 386 AA.
AC Q9Y5R3	01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999	(TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand	
DE sulfotransferase GST-3).	
CN GST3.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	SEQUENCE FROM N.A.
RP	SEQUENCE FROM N.A.
RC TISSUE=Tonsil;	
RX MEDLINE=99264336; PubMed=10330415;	
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,	
RT Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;	
RT "Sulfotransferases of two specificities function in the reconstitution	
RT of high endothelial cell ligands for L-selectin.";	
RL J. Cell Biol. 145:899-910(1999).	
RN [2]	SEQUENCE FROM N.A.
RP	SEQUENCE FROM N.A.

Q8NCG5 homo sapien
Q8IV46 homo sapien
Q9RL11 mus musculus
Q9WUE5 mus musculus
Q9GXX3 homo sapien
Q9GXP4 mus sapien
Q9GZS9 homo sapien
Q9EP78 mus musculus
Q99NB0 mus musculus
Q75667 homo sapien
Q9NE84 homo sapien
Q8N876 mus musculus
Q80W73 mus musculus
Q9UED5 homo sapien
Q9V4C5 homo sapien
Q9EQC0 mus musculus
Q43916 homo sapien
Q88199 mus musculus
Q75099 homo sapien
Q9QZ12 rattus norv
Q93403 torpedo cal
Q9VNC3 drosophila
Q9DOK5 mus musculus
Q9VNC4 drosophila
Q8MZD1 drosophila
Q95J48 cryptotagus
Q75851 homo sapien
Q82V94 rhizobium m
Q8UZE1 cercopithec
Q68872 myxococcus
Q8SPM4 bos taurus
Q13896 homo sapien
Q810J9 mus musculus
Q76045 homo sapien
Q8N473 homo sapien
Q9TTS5 bos taurus
Q8GFF2 streptomyce
Q25001 globodera p
Q94438 chironomus
Q9GK59 herpesvirus
Q8CG65 mus musculus
Q8SPM4 bos taurus
O55000 rattus norv
Q86XJ2 homo sapien

Qy	1028	CATCTTCAGACCTGGGTGCATAACATCACCGAGGCAAGCGCATGGTGACACGCTTTC	108
Db	301	HisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaIle	320
Qy	1088	CACACAAATGCCAGGGATGCCCTAAATGCTCTCCAGGCTTGGCGCTGGTCTTTGCCCTAT	1147
Db	321	HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTyr	340
Qy	1148	GAAGAAGTTTCTCGACTTCAGAAAGCCTGGCGATGCCATGAATTTGCTGGGCTACCGC	1207
Db	341	GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrArg	360
Qy	1208	CACGTCAGATCTGAACAAGAACAACAGAAAACCTGTTGCTGGATCTTCTCTACCTGGACT	1267
Db	361	HisValArgSerGluGlnGluArgAsnLeuLeuAspLeuLeuSerThrTrpThr	380
Qy	1268	GTCCTCGAGCAAAATCCAC	1285
Db	381	ValProGluGlnIleHis	386
RESULT 3			
Qy	Q8IV46	PRELIMINARY; PRT; 370 AA.	
Db	Q8IV46;		
AC	Q8IV46;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Similar to carbohydrate (N-acetylglucosamine 6-O) sulfotransferase		
DE	4.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_taxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	IISUB=Brain;		
RA	Strausberg R.;		
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC035282; AAH35282.1; -		
DR	GO; GO:0008146; F:sulfotransferase activity; IEA.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	InterPro; IPR001092; HLH basic.		
DR	InterPro; IPR000863; Sulfotransferase.		
DR	Pfam; PF00685; Sulfotransfer; 1.		
DR	PROSITE; PS00038; HLH_1; 1.		
KW	Transferase.		
SQ	SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;		
Alignment Scores:			
Pred. No.:	4,33e-160	Length:	370
Score:	1963.00	Matches:	370
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	52.46%	Indels:	0
DB:	4	Gaps:	0
US-09-645-078-1 (1-2043) x Q8IV46 (1-370)			
Qy	176	ATGCCCATCTTGGCTCTATTCTTCCACATGTACGCCAACATCAGTCCCTGCTATG	235
Db	1	MetAlaIleLeuAlaLeuPheHisMetTyrSerHisAsnIleSerSerLeuSerMet	20
Qy	236	AAGCAGACCCCGAGCGATGATGATGCTGCTTCTTCTGCTGCTGCTGCTGCTTCT	295
Db	21	LysAlaGlnProGluArgMetHisValLeuValLeuSerSerTrpArgSerGlySerSer	40
Qy	296	TTTGTGGGGCAGCTTTTGGGCGAGCACCACAGATGTTTCTACCTGATGGAGCCCGCTGG	355
Db	41	PheValGlyGlnLeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrp	60
Qy	356	CACGTGTGGATGACCTTCAAGCAGACGACCGCTGGATGCTGACATGCTGCGGGAT	415

Db	61	HisValTrpMetThrPheLysGlnSerThrAlaTrpMetLeuHisMetAlaValArgAsp 80
Qy	416	CTGATAGCGCGCTCTTGTGGGACATGAGCGTCTTTCATCCCTACATGGAACCTGTT 475
Db	81	LeuileArgAlaValPheLeuLysCysAspMetSerValPheAspAlaTrpMetGluProGly 100
Qy	476	CCCCGGGAGACAGTCCAGCCTCTTTCAGTGGGAGAACACCGCGCCCTGTCTTTCACCT 535
Db	101	ProArgGlnSerSerLeuPheGlnTrpGluAsnSerArgAlaLeuLysSerAlaPro 120
Qy	536	CCCTGTGACATCCCAAGATGAATCATCCCGCGCTCCTCAGCTCAGCGCTCCGTGTC 595
Db	121	AlaCysAspIleIleProGlnAspGluIleIleProArgAlaHisCysArgLeuLeuLys 140
Qy	596	AGTCAACAGCCCTTGTAGCTGGTGGAGAGCCCTCCGCTCTACAGCCACGTGGTCTC 655
Db	141	SerGlnGlnProPheGluValValGluLysAlaCysArgSerTyrSerHisValValLeu 160
Qy	656	AAGAGGTGCGCTTCTTCAACTGACGTCCCTCTACCGCTGCTGAAAGACCCCTCCCTC 715
Db	161	LysGluValArgPheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeu 180
Qy	716	AACCTGATATCGTCACCTGTCGGGACCCCGCGCTGTTCCGTTCCTCCGAGAACGC 775
Db	181	AsnLeuHisIleValHisLeuValArgAspProArgAlaValPheArgSerArgGluArg 200
Qy	776	ACAAAGGAGATCTCATGATTCACAGTCGCGATTGTGATGGGCGAGCATGAGCAAACTC 835
Db	201	ThrLysGlyAspLeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeu 220
Qy	836	RAGAGGAGGACCACTCTATGATGATGAGGTCACTGCGCAAGCCACTCGAGATC 895
Db	221	LysLysGluAspGlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIle 240
Qy	896	TACAAGACCATCCAGCTCTTCCCAAGCCCTGCGAGAACCTACCTCTTGTGCGCTAT 955
Db	241	TyrLysThrIleGlnSerLeuProLysAlaLeuGlnGluArgTyrLeuValArgTyr 260
Qy	956	GAGGACCTGCTGAGCCCTGTCGGCCAGCTTCCGAAATGATGAATTCGTGGGATG 1015
Db	261	GluAspLeuAlaArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeu 280
Qy	1016	GAATTCCTCCCATCTTCAGACCTGGGTGCATAACATCACCGGAGCAAGGCGATGGT 1075
Db	281	GluPheLeuProHisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGly 300
Qy	1076	GACCAGCTTTCACAAATGCCAGGATGCCCTTAATGTCTCCAGCTTGGCGCTGG 1135
Db	301	AspHisAlaPheHisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrp 320
Qy	1136	TCTTTGCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCCTGTGGCGATGCCATGATTG 1195
Db	321	SerLeuProTyrGluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeu 340
Qy	1196	CTGGCTACCGCCACGTCTGATCTGAACAGAACAGAGAACTGTCTGCTGATCTTCTG 1255
Db	341	LeuGlyTyrArgHisValArgSerGluGlnGluGlnArgAsnLeuLeuAspLeuLeu 360
Qy	1256	TCTACCTGGACTGCTCCTCAGCAAAATCCAC 1285
Db	361	SerThrTrpThrValProGluGlnIleHis 370
RESULT 4		
ID	Q9R111	PRELIMINARY; PRT; 388 AA.
AC	Q9R111	
DT	01-MAY-2000	(TrEMBLrel. 13, Created)
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)
DE	L-selectin ligand sulfotransferase.	
GN	CHST4.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99361934; PubMed=10435581;
RA	Hirooka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
RA	Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.,
RT	"A novel, high endothelial venule-specific sulfotransferase expresses
RT	6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.",
RL	Immunity 11:79-89(1999).
DR	EMBL; AF109155; AAD45579.1; -.
DR	MGD; MGI:1349479; Chst4.
DR	GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR	GO; GO:0005529; F:sugar binding; IEA.
DR	GO; GO:0008146; F:sulfotransferase activity; IEA.
DR	GO; GO:0016740; F:transferase activity; IEA.
DR	GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR	InterPro; IPR000863; Sulfotransferase.
DR	Pfam; PF00685; Sulfotransfer; 1.
KW	Lectin; Selectin; Transferase.
SQ	SEQUENCE 388 AA; 44636 MW; 6D5371APB6884AEE CRC64;
Alignment Scores:	
Pred. No.:	4,548-122 Length: 388
Score:	1521.50 Matches: 283
Percent Similarity:	85.13% Conservative: 49
Best Local Similarity:	72.56% Mismatches: 55
Query Match:	40.66% Indels: 3
DB:	11 Gaps: 2
US-09-645-078-1 (1-2043) x Q9R111 (1-388)	
Qy	128 ATGCTACTGCCTAAATAAATGAAGCTCTGCTGTTCTTGTGTTCCAGATGCCATCTTG 187
Db	1 MetMetLeuLysLysGlyArgLeuLeuMetPheLeuGlySerGlnValIleValVal 20
Qy	188 GCTCTATTCTTCACATGTACAGCCACACATCAGCTCCCTCTCTATGAGCACAGCC 247
Db	21 AlaLeuPheIleHisMetSerValHisArg-----HisLeuSerGlnArgGluGluSer 38
Qy	248 GACGCG---ATGCACCTGCTGCTGCTCTCTCTGCGCTCTGCTCTCTTTTGTGGG 304
Db	39 ArgArgProValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGly 58
Qy	305 CAGCTTTTGGGAGACACCCAGATGTTTCTACCTGATGAGCCCGCTGGCAGCTGTGG 364
Db	59 GlnLeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrp 78
Qy	365 ATGACCTTCAAGCAGACGACCGCTCGATGCTGCACATGCTGCGGATCTGATACGG 424
Db	79 MetThrPheThrSerThrAlaTrpLysLeuHisMetAlaValArgAspLeuLeuArg 98
Qy	425 GCCCTCTTCTTGTGGACATGAGCGTCTTTGATGCTTACATGGAACCTGTCCTCCG 484
Db	99 SerValPheLeuLysCysAspMetSerValPheAspAlaTyrMetAsnProGlyProArgLys 118
Qy	485 CAGTCCAGCTCTTTCAGTGGGAGAACAGCCGCGCTGTTCTGCTCTGCTGCTGTGAC 544
Db	119 GlnSerSerLeuPheGlnTrpGluGlnSerArgAlaLeuLysSerAlaProValCysAsp 138
Qy	545 ATCATCCCAACAGATGAATCATCCCGCGGCTCCTGTCAGGCTCTCTGTGCTGACAG 604
Db	139 PhePheProAlaHisGluIleSerSerProLysHisCysLysLeuLeuLysGlnGln 158
Qy	605 CCCTTTGAGGTGGGAGAGCGCTGCTGCTCTACAGCCACGCTGCTCAAGGAGGTG 664
Db	159 ProPheAspMetValGluLysAlaCysArgSerHisGlyPheValValLeuLysGluVal 178
Qy	665 GCCTTCTTCAACCTGAGTCCCTCTACCCGCTGCTGTAAGACCCCTCCCTCAACCTGCAT 724
Db	179 ArgPheLeuSerLeuGlnAlaLeuTyrProLeuLeuThrAspProSerLeuAsnLeuHis 198
Qy	725 ATCGTGACCTGTTCCGGGACCCCGCGCGCTGTTCCGTTCCCGAGAACCCAGAGGGA 784

```

Db 199 ValValHisLeuValA-gsppProArgAlaValPheArgSerArgGluHisThrThrile 218
QY 785 GATCTCATGATTGACAGTCGCGATTGTGATGGGCGAGCATGAGCAAAAACCTCAAGAGGAG 844
Db 219 GluLeuValValAspSerHisIleValLeuGlyGlnHisLeuGluThrIleLysGluGlu 238
QY 845 GACCAACCTACTATGTGATGACGGTANTGCGCAAGCCAGCTGGAGATCTACAGACC 904
Db 239 AspGlnProTyrTyraAlaMetLysIleleCysLysSerGlnValAspIleValLysAla 258
QY 905 ATCCAGTCTTGGCCCAAGCCCTGAGGAACGCTACCTGCTGTGCGCTATGAGGACCTG 964
Db 259 IleGlnThrLeuProGluAlaLeuGlnGlnArgTyrLeuPheLeuA-gTyrGluAspLeu 278
QY 965 GCTGAGCCCTGTGGCCGACACTCCCGAATGATGAATTCGTGGGATTGGAAATCTTG 1024
Db 279 ValArgAlaProLeuAlaGlnThrThrArgLeuTyrLysPheValGlyLeuAspPheLeu 298
QY 1025 CCCATCTTCAGACTGGTGCATAAACATCACCGAGCGAGGCGATGGGTGACACGCT 1084
Db 299 ProHisLeuGlnThrTrpValHisAsnValThrArgGlyLysGlyMetGlyGlnHisAla 318
QY 1085 TTCCACAAATGCCAGGAGCCCTTAATCTCTCCAGGCTTGGCGCTGCTTTTCCCC 1144
Db 319 PheHisThrAsnAlaArgAsnAlaLeuAsnValSerGlnAlaTrpA-gTrpSerLeuPro 338
QY 1145 TATGAAGGTTTCTCGACTTCAGAAACCTGTGGCGATGCGATGAATTCCTGGGCTAC 1204
Db 339 TyrGluLysValSerGlnLeuGlnAspAlaCysGlyGlnAlaMetAspLeuLeuGlyTyr 358
QY 1205 CGCCAGCTCAGATCTGAACAAGACAGAGAACCTGTGCTGGATCTTCTCTACCTGG 1264
Db 359 LeuGlnValArgSerGlnGlnGlnGlyAsnLeuSerLeuAspLeuLeuSerSer 378
QY 1265 ACTGTCCTGAGCAATCCACTAAGAGGT 1294
Db 379 HisIleLeuGlyGlnValPheArgGluGly 388

RESULT 5
Q9WUE5
ID Q9WUE5 PRELIMINARY; PRT; 388 AA.
AC Q9WUE5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin
DE 6/keratan) sulfotransferase 4).
GN CHS74.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,
RA Huang C.-C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

```

```

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RL EMBL; AF131236; AAD33016.1; -.
DR EMBL; AK009113; BAB26078.1; -.
DR MGD; MGI:1349479; Chr24.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

```

```

Alignment Scores:
Pred. No.: 6,75e-122 Length: 388
Score: 1519.50 Matches: 283
Percent Similarity: 85.13% Conservative: 49
Best Local Similarity: 72.56% Mismatches: 55
Query Match: 40.61% Indels: 3
DB: 11 Gaps: 2

```

US-09-645-078-1 (1-2043) x Q9WUE5 (1-388)

```

QY 128 ATGCTACTCCCTAAATAAATGAAGCTCTCTGCTGTTCTTGTGTTTCCAGATGCCATCTG 187
Db 1 MetMetLeuLeuLysLysGlyArgLeuLeuMetPheLeuGlySerGlnValVal 20
QY 188 GCTCTATTCTTCCCATGTACAGCCACACATCAGCTCCCTGCTCTATGAGGCACAGCC 247
Db 21 AlaLeuPheIleHisMetSerValHisArg-----HisLeuSerGlnArgGluGluSer 38
QY 248 GAGCGC---ATGCACGCTGCTGCTTCTGCTTCTTCTGCGCTCTGGCTTCTTTTGTGGG 304
Db 39 ArgArgProValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGly 58
QY 305 CAGCTTTTGGGAGCAGCCAGCATGTTTCTACCTGATGGAGCCCGCTGGCAGCTGGG 364
Db 59 GlnLeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrp 78
QY 365 ATGACCTTCAAGCAGACGCGCTGATGCTGCATGCTGCGGATCTGATACCG 424
Db 79 MetThrPheThrSerSerThrAlaTrpLysLeuHisMetAlaValArgAspLeuLeuArg 98
QY 425 GCGCTTCTTGTGCCACATGAGCGCTTTTGATGCTTACATGGAACCTGCTGCTCCCGGAGA 484
Db 99 SerValPheLeuCysAspMetSerValPheAspAlaTyrMetAsnProGlyProArgLys 118
QY 485 CAGTCCAGCTCTTTCAGTGGGAGACGCGCGCTGTTCTGCTGCTGCTGCTGCTGCTGAC 544
Db 119 GlnSerSerLeuPheGlnTrpGluGlnSerArgAlaLeuCysSerAlaProValCysAsp 138
QY 545 ATCATCCCAACAGATGAATCATCCCGCGCTCTACTGAGGCTCTGTCAGTCAACAG 604
Db 139 PhePheProAlaHisGluIleSerSerProLysHisCysLysLeuLeuCysGlyGlnGln 158
QY 605 CCCTTTGAGGTGGTGAAGAAGGCTCCGCTCTCTACAGCCAGTGGTGGTCTCAGGAGGTG 664
Db 159 ProPheAspMetValGluLysAlaCysArgSerHisGlyPheValValLeuLysGluVal 178
QY 665 CGCTTCTTCAACTGCGAGTCCCTCTACCCGCTCTGCTGAAAAGACCCCTCCCTCAACCTGCAT 724
Db 179 ArgPheLeuSerLeuGlnAlaLeuTyrProLeuLeuThrAspProSerLeuAsnLeuHis 198

```



```
Db 256 AlathrLeuLysPro---ProPheLeuArgGlyArgTyrArgLeuValArgPheGlu 274
QY 959 GACCTGGCTCCAGCCCTGTGGCCAGACTTCCCGAATGATGAATTCGTGGGATTGGAA 1018
Db 275 AspLeuAlaArgGluProLeuAlaGluLeuArgAlaLeuTyrAlaPheThrGlyLeuSer 294
QY 1019 TTCTGGCCCAATCTCAGACCTGGGTGCATTAACATCACCCGAGCAAGGGCATGGGT--- 1075
Db 295 LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 --GACCACGCTTCCACACAAATGCCAGGATGCCCTTAATGCTCCAGGCTTGGCGC 1132
Db 315 ArgArgGluAlaPheLeuThrSerSerArgAsnAlaLeuValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTGGCCCTAAGAAAGTTCTCGACTTCAGAAAGCTGTGGCGATGCCATGAAT 1192
Db 335 HisAlaLeuProPheAlaLysIleArgArgValGlnGluLeuCysAlaGlyAlaLeuGln 354
QY 1193 TTGTGGGCTACCGCAGCTCAGATCTCAACAAGACAGAGAAACCTGTTGCTGATCTT 1252
Db 355 LeuLeuGlyTyrArgProValTyrSerGluAspGlnArgAsnLeuAlaLeuAspLeu 374
QY 1253 CTGTCTACCTGGACTGTCCCTCAGCAATCCATCAAGAGGTTGAGAAGGCTTTGCTGCC 1312
Db 375 Val-----LeuPro-----ArgGlyLeuAsnGlyPhe----- 383
QY 1313 ACCTGGTGTACGCTCAGTCACTTCTCTGAATGCTTCTGAGCCT 1357
Db 384 ThrTrpAla-----SerSerThrAlaSerHisPro 393

RESULT 7
Q9QUP4 PRELIMINARY; PRT; 395 AA.
AC Q9QUP4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN CHST5 OR 1-GLCNAC-6-ST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Intestine;
RX MEDLINE=99423499; PubMed=10491328;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
RL Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
DR EMBL, AF176841; AAD56003.1; -;
DR EMBL, AF176840; AAD56002.1; -;
DR MGD, MGI:1931825; Chst5.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 395 AA; 44537 MW; - 3PDF71E43ED383BE CRC64;

Alignment Scores:
Pred. No.: 6,92e-79 Length: 395
Score: 1020.00 Matches: 221
Percent Similarity: 66.58% Conservative: 44
Best Local Similarity: 55.53% Mismatches: 120
Query Match: 27, 26% Indels: 13
DB: 11 Gaps: 8

US-09-645-078-1 (1-2043) x Q9QUP4 (1-395)
QY 128 ATGCTACTGCCTAAA-----AAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGCC 181
```

```
Db 1 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuLeuMetValGlnThrGly 20
QY 182 ATCTTGGCTCTATTCTTCCACATGTACAGCCACACATCAGCTCCCTCTCTATGAAGCA 241
Db 21 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 36
QY 242 CAGCCCGAGCGCATCACTGCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 301
Db 37 LeuGlyGluArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheVal 56
QY 302 GGGCAGCTTTTGGGCGAGCACCCAGATGTTTCTACCTGATGGAGCCCGCTGGCACGTG 361
Db 57 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal 76
QY 362 TGGATGACCTTCAAGCAGACACCCCTGGATGCTGTCACATGGCTGCGGGATCTGATA 421
Db 77 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeuIle 96
QY 422 CGGGCCGCTCTTCTGTGCGACATGAGCGTCTTTGATGCTTACATCGAAACCTGGTCCCGG 481
Db 97 ArgSerValPheLeuCysAspMetAspValPheAspAlaTyrLeu---ProTrpArgArg 115
QY 482 AGACAGTCAGCTCTTTCAGTGGGAGAACAGCCGGGCTGTGTCTGTCACCTGCCTGT 541
Db 116 AsnIleSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProValCys 135
QY 542 GACATCATCCCAAGATGAATCATCCCGGGCTCACTGCAGCTCTCTGTCAGTCAA 601
Db 136 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLysProLeuCysAlaThr 155
QY 602 CAGCCCTTTGAGTGGAGAAAGCCCTGCTCCTACAGCCACGTGGTCTCAAGAG 661
Db 156 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 175
QY 662 GTGCGCTTCTTCAACTGCACTCCCTTACCCGCTGCTGAAAGACCCCTCCCTCAACCTG 721
Db 176 ValArgPheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 195
QY 722 CATATCGTCACCTGTGCGGACCCCGGGCGCTGTTCCGTTCCCGAAGACGCAACAAG 781
Db 196 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAla 215
QY 782 GAGATCTCATGATTCAGTCCGCTTGTGATGGGGCAGCATGACCAAAACCTCAAGAAG 841
Db 216 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---Glu 234
QY 842 GAGGACCAACCTTACTATGTGTGATGAGGTCTATGCCAAAGCCAGCTGGAGATCTACAAG 901
Db 235 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 254
QY 902 ---ACCATCAAGTCTTGGCCCAAGGCCCTGCGAGGAACGCTACTGCTGTGGCTATGAG 958
Db 255 AlaAlaLeuHisLysProProPheLeuGlnAspArgTyrArgLeuValArgTyrGlu 274
QY 959 GACCTGCTCGAGCCCTGTGGCCAGACTTCCCGAATGTATGAATTCGTGGATTGGAA 1018
Db 275 AspLeuAlaArgAspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGly 294
QY 1019 TTCTTGGCCCTTCCAGACTGGGTGCATAACATCACCCGAGCAAGGGCATGGGT--- 1075
Db 295 LeuThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCACGCTTCCACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTTGGCGC 1132
Db 315 ArgArgGluAlaPheLysThrThrSerArgAspAlaLeuSerValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTGGCCCTATGAAAAGGTTTCTCGACTTCAGAAAGCTGTGGCGATGCCATGAAT 1192
Db 335 HisThrLeuProPheAlaLysIleArgArgValGlnGluLeuCysGlyGlyAlaLeuGln 354
QY 1193 TTGCTGGGCTACCGCAGCTCAGATCTGAACAGAGAACCTGTGCTGATCTT 1252
```



```

QY 913 CTGCCCCAGCCCTGTCAGGAAGCTACCTGCTGTGTGCGCTATGAGGACCTGGCTCGAGC 972
Db      |||
QY 359 YAlaProAlaTrieuArgArgTyrLeuArgLeuArgTyrGluAspLeuValTrieuG1 379
Db      |||
QY 973 CCCTGTGGCCAGACCTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1032
Db      |||
QY 379 nProGlnAlaGlnLeuArgLeuArgLeuArgLeuArgLeuArgLeuArgLeuArgLeuArg 399
QY 1033 TCAGACCTGGTGCATACATACCCGAGGACGAGGATGGGT--GACACGCTTTCCA 1089
Db      |||
QY 399 uAspAlaPheAlaPheAsnMetThrArgGlySerAlaTyrGlyAlaAspArgProPheHi 419
QY 1090 CACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTGGCTGGCTGGCTGGCTGGCTGGCT 1149
Db      |||
QY 419 sLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrieuArgLeuSerGlnG1 439
QY 1150 AAAGTTCTTCGATCTCAGAAACCTGTGCGCATGCCATGAATTTGCTGGGTACCGCCA 1209
Db      |||
QY 439 uGlnValArgGlnValGluThrAlaCysAlaProAlaMetArgLeuLeuAlaTyrProAr 459
QY 1210 CGTCAGATCTGAACAAGACAGAGAAACCTGTG-----CTGATCTTCT 1254
Db      |||
QY 459 gSerGlyAspGluArgAspArgLysThrValArgGluGlyGluThrProLeuGluThrLy 479
QY 1255 GTCTACCTGGACTGTC 1270
Db      |||
QY 479 sAlaAsnTrieuAlaVal 484

RESULT 10
Q99NB0 PRELIMINARY; PRT; 484 AA.
AC Q99NB0;
AT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Chondroitin 6-sulfotransferase-2.
EN CHST7 OR MC6ST-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
FX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Kitagawa H., Uyama T., Sugahara K.;
RT "Cloning and Expression of Mouse Chondroitin 6-sulfotransferase-2."
EL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046929; BAB40372.1; -.
DR PIR; JC7350; JC7350.
DR MGD; MGI:1891767; Chst7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 484 AA; 54751 MW; 402C5E1ED185DF8 CRC64;

Alignment Scores:
Pred. No.: 3,09e-45 Length: 484
Score: 629.50 Matches: 154
Percent Similarity: 48.50% Conservative: 72
Best Local Similarity: 33.05% Mismatches: 176
Query Match: 16.82% Indels: 64
DB: 11 Gaps: 11

US-09-645-078-1 (1-2043) x Q99NB0 (1-484)
QY 19 TACGGAAGACGACAGAGGGTAGAGGAGAAAGCGATGGCCCGG---CTAGCAGTG--- 72
Db      |||
QY 35 HisserGluGlnAspLysGlyArgAsnCysProGlyLeuGlnArgSerLeuGlyValTrieu 54
QY 73 AGCTCTCAAAAGCAGAGGAGGAGCCCAAGCCACAGGCTCTCCACTTCAGACACATGCT 132
Db      |||

```

```

Db 55 SerLeuGluAlaAlaAlaGluArgGluGlnGlyAlaGluValArgSerLeuAla 74
QY 133 ACTGCGCTAAATAAATGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 192
Db      |||
QY 75 -----GluGlyAsnProAspArgSerProGlySer 84
QY 193 ATTCTTCCACATGTACACCCACAAACATCAGCTCCCTGTCTATGAGGACGACAGCCCGAGCG 252
Db      |||
QY 85 -----Pro-GlyAsnLeuSerAlaValGlyGluAlaValThrGlnGluLy 99
QY 253 CATGACAGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 312
Db      |||
QY 99 sGlnHisIleTyrValHisAlaThrTrieuArgThrGlySerSerPheLeuGlyGluLeuPh 119
QY 313 TGGGAGACACCCAGATGTTTCTACCTGATGAGAGCCCGCTGGCCACGTGTGGATGACCTT 372
Db      |||
QY 119 eAsnGlnHisProAspValPheTyrLeuTyrAspProMetTrieuHisTrieuGlnAlaLe 139
QY 373 CAAGCAGACACCGCTCGATGCTGCACATGGCTGTGGGATCTGATAGCGGCCGCTCT 432
Db      |||
QY 139 uTyrProGlyAspAlaGluSerLeuGlnGlyAlaLeuArgAspMetLeuArgSerLeuPh 159
QY 433 CTGTGCGACATGAGCGTCTTGTGCTACATGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 487
Db      |||
QY 159 eArgCysAspPheSerValLeuArgLeuTyrAlaGlnProGlyAspProGlyGluArgAl 179
QY 486 -----TCCAGCGCTCTTTTCAGTGGGAGAACACGCGCGCCCT 522
Db      |||
QY 179 aProAspSerAlaAsnLeuThrAlaMetLeuPheArgTrieuArgThrAsnLysValI1 199
QY 523 GTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 570
Db      |||
QY 199 eCysSerProProLeuCysProAlaAlaProArgAlaAlaArgAlaAspValGlyLeuValG1 219
QY 571 CCGGGCTCACTGCGGCTCTGTGCGTCAACAGCCCTTTGAGGTGGTGGAGAGGCGCTG 630
Db      |||
QY 219 uAspLysAlaCysGluSerThrCysProProValSerLeuArgAlaLeuGluAlaGluCy 239
QY 631 CGCTCTTACACCCAGCTGTGCTCAAGAGGTGGCTTCTTCACTGAGCTGCTGCTTCA 690
Db      |||
QY 239 sArgLysTyrProValValIleLysAspValArgLeuAspLeuGlyValLeuVa 259
QY 691 CCGCTGTCTGAAAGACCCCTCCCTCAACTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 750
Db      |||
QY 259 lProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheArgAspProAr 279
QY 751 GCGCGTGTCTCTCCCGAGAACCGCAAGGAGATCTCATGATTGACAGT----- 802
Db      |||
QY 279 gAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgLuserIleGlnVa 299
QY 803 -----CGATTCTGATG----- 814
QY 299 lLeuArgThrArgGlnArgGlyAspHisPheHisArgValLeuLeuAlaHisGlyValAs 319
QY 815 -----GGGACGCTGAGCAAAACTCAAGAGGAGGACCAACCTACTAT----- 859
QY 319 pAlaArgProGlyGlyGlnAlaArgAlaLeuProSerAlaProArgAlaAspPhePheLe 339
QY 860 -----GTGATGACAGTCTCTCCAAAGCCAGCTGAGATCTACAGACCATCCCATC 912
Db      |||
QY 339 uThrSerAlaLeuGluValIleCysGluAlaTrieuArgAspLeuLeuPheThrArgG1 359
QY 913 CTGCCCCAGGCGCTGCGAGAACGCTACTCTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 972
Db      |||
QY 359 yAlaProAlaTrieuArgArgTyrLeuArgLeuArgTyrGluAspLeuValTrieuG1 379
QY 973 CCCTGTGGCCAGACCTCCCGAATGTATGAAATTCGTGGATTTGGAATTTCTTGGCCCATCT 1032
Db      |||
QY 379 nProGlnAlaGlnLeuArgArgLeuArgPheSerGlyLeuArgThrLeuAlaAlaLe 399
QY 1033 TCAGACCTGGTGCATACATCACCAGGAGGAGGATGGGT---GACACGCTTTCCA 1089
Db      |||
QY 399 uAspAlaPheAlaPheAsnMetThrArgGlySerAlaTyrGlyAlaAspArgProPheHi 419

```

```

QY 1090 CACAAATCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGGTCTTGGCCCTATGA 1149
Db 419 sLeuSerAlaArgAspAlaA:GluAlaValHisAlaTTPArgGluArgLeuSerGlnGI 439
QY 1150 AAAGCTTCTCGACTTCAGAAAGCTGTGGCGATGCCATGAATTGCTGGGCTACCCGCA 1209
Db 439 uGlnValArgGlnValGluThrAlaCysAlaProAlaMetArgLeuAlaTyrProAr 459
QY 1210 CGTCAGATCTGAACAAGAACAGAGAAACCTGTG-----CTGGATCTCTCT 1254
Db 459 gSerGlyAspGluArgAspArgLysThrValArgGluGluThrProLeuGluThrLy 479
QY 1255 GTCTACCTGGAGTCTC 1270
Db 479 sAlaAsnTrpAlaVal 484

RESULT 11
O75667 PRELIMINARY; PRT; 486 AA.
ID AC O75667;
DC O75667;
AT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (Carbohydrate
DE (N-acetylglucosamine 6-O) sulfotransferase 7).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Uchimura K., Muramatsu T.;
RT Identification and molecular characterization of a cDNA encoding a
RT novel N-acetylglucosamine-6-O-sulfotransferase."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB040711; BAB13770.1; -;
DR EMBL; BC045537; AAH45537.1; -;
DR PIR; JC7351; JC7351.
DR Genew; HGNC:13817; CHST7.
DR GO; GO:0008145; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 486 AA; 54266 MW; 3F1FD1430B3C8E95 CRC64;

Alignment Scores:
Pred. No.: 4,97e-44 Length: 486
Score: 615.50 Matches: 155
Percent Similarity: 48.22% Conservative: 62
Best Local Similarity: 34.44% Mismatches: 163
Query Match: 16.45% Indels: 70
DB: 4 Gaps: 12

US-09-645-078-1 (1-2043) x O75667 (1-486)
QY 7 CATTCTGTGGTACGGAGACGACAGAGGGTAGAGGAGAAAGCGCATGGCCCGCTA 66
Db 44 HisCysProGlyLeuGlnArg-----SerLeu 52
QY 67 GCAGTG---AGCCTCTCAAAAGCAGCAGGAGAGCCCAAGCCCAAGGCTTCCACTTCAG 123
Db 53 GlyValTrpSerLeuGluAlaAlaAlaGlyGluArgGluGluGlnGlyAlaGluAlaArg 72
QY 124 CACAATGCTACTGCTCTAAATAAAGTCTCTGCTTTCTGTTTCCAGATGGCCAT 183

```

```

Db 73 AlaAlaGluGluGlyGlyAlaAsnGlnSerProArg-----PheProSerAsnLeu 89
QY 184 CTTGGCTCTATTCTTCCACATGTACAGCCACACATCAGCTCCCTGTCTATGAAGGCACA 243
Db 90 SerGlyAlaValGlyGluAlaVal-SerArg----- 99
QY 244 GCCCGAGCCGATGCAGGCTGCTGTCTCTCTCTGGCGCTGCTGCTCTCTCTTTTGTGG 303
Db 100 ----GluLysGlnHisIleTyrValHisAlaThrTrpArgThrGlySerSerPheLeuGI 118
QY 304 GCAGCTTTTGGCGAGCACCAGATGTTTCTACCTGATGGAGCCCGCTGCGCATGTG 363
Db 118 yGluLeuPheAsnGlnHisProAspValPheTyrLeuTyrGluProMetTrpHisLeuTr 138
QY 364 GATGACCTTCAAGCAGAGACCCGCTGGATGCTGCATGCTGGTGTGGGATCTCATAG 423
Db 138 pGlnAlaLeuTyrProGlyAspAlaLeuSerLeuGlnGlyAlaLeuArgAspMetLeuAr 158
QY 424 GGCGCTCTTCTTGTGGACATGAGGCTCTTTGATGCCTACATGGAACCTGGT---CCCG 480
Db 158 gSerLeuPheArgCysAspPheSerValLeuArgLeuTyrAlaProProGlyAspProAl 178
QY 481 GAGACAG-----TCCAGCCTCTTTTCAGTGGGAGAACAG 513
Db 178 aAlaArgAlaProAspThrAlaAsnLeuThrThrAlaAlaLeuPheArgTrpArgThra 198
QY 514 CGGGCCCTGTGTCTGCACCTGCTGTGACATCATCCACAGATGAA----- 562
Db 198 nLysValIleCysSerProProLeuCysProGlyAlaProArgAlaArgAlaGluValGI 218
QY 563 -ATCATCCCCCGGCTCACTGCAGGCTCTGTGCAGTCAACAGCCCTTTGAGTGGTGA 621
Db 218 yLeuValGluAspThrAlaCysGluArgSerCysProValAlaIleArgAlaLeuGI 238
QY 622 GAAGGCTCCCTCTCTACAGCAGCTGTGTCTCAAGAGGTGCGCTTCTTCAACTGCA 681
Db 238 uAlaGluCysArgLysTyrProValValIleLysAspValArgLeuLeuAspLeuGI 258
QY 682 GTCCTCTACCCGCTGCTGAAGACCCCTCCCTCACTGCATATCGTCACCTGGTCCG 741
Db 258 yValLeuValProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheAr 278
QY 742 GAGCCCGCGCGCTTCCGTTCCGAGAACGCGACAAAGGGAGATCTCATGATTGACAG 801
Db 278 gAspProArgAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgLysLe 298
QY 802 T-----CGCATTTGTGATG----- 814
Db 298 rIleGlnValLeuArgThrArgGlnArgGlyAspArgPheHisArgValLeuLeuAlaHi 318
QY 815 -----GGCGAGCATGAGCAAAAACCTCAAGAGGAGGACCAACCCTA 855
Db 318 sGlyValGlyAlaArgProGlyGlyGlnSerArgAlaLeuProAlaAlaProArgAlaAs 338
QY 856 CTAT-----GTGATGCGGTCTATCGCCAAAGCCAGCTGGAGATCTACAAGAC 903
Db 338 pPhePheLeuThrGlyAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuLeuPh 358
QY 904 CATCCAGTCTTGGCCCAAGCCCTGCAGGAACGCTACTGCTGTGTGGCTATGAGGACCT 963
Db 358 eAlaArgGlyAlaProAlaTrpLeuArgArgTyrLeuArgLeuLeuArgTyrGluAspLe 378
QY 964 GGCTCGAGCCCTGTGGCCCGACACTTCCCGATGTGAATTCGTGGGATTCGAATTCCT 1023
Db 378 uValArgGlnProArgAlaGlnLeuArgLeuLeuArgPheSerGlyLeuArgAlaLe 398
QY 1024 GCCCCATCTTCAGACTGGGTGCATACATACCCGAGCAGCAAGGCGATGGGT---GACCA 1080
Db 398 uAlaAlaLeuAspAlaPheAlaLeuAsnMetThrArgGlyAlaAlaTyrGlyAlaAspAr 418
QY 1081 CGCTTTCCACAAATGCGGAGGATGCCCTTAATGCTCTCCAGGCTTGGCGTGGTCTTT 1140
Db 418 gProPheHisLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrpArgGluArgLe 438

```

QY 1141 GCCTATGAAAGTTCTCGACTTCAGAAAGCCTGTGGCGATGCCATGTAATTTGCTGGG 1200
Db 438 userArgGluGlnValArgGlnValGluAlaCysAlaProAlaMetArgLeuLeuAl 458
QY 1201 CTACGCCACGTCAGATCTGAACAGAA 1228
Db 458 atyr-----ProArgSerGlyGluGlu 465

RESULT 12
Q9NS84 PRELIMINARY; PRT; 486 AA.
AC Q9NS84;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Chondroitin 6-sulfotransferase-2.
GN C6ST-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20347214; PubMed=10781596;
RA Kitagawa H., Fujita M., Ito N., Sugahara K.;
RT "Molecular cloning and expression of a novel chondroitin 6-O-
sulfotransferase.";
RL J. Biol. Chem. 275:21075-21080(2000).
DR EMBL; AB037187; BAB03217.1; .
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0008459; F:chondroitin 6-sulfotransferase activity; TAS.
DR GO; GO:0005976; P:polyaccharide metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 486 AA; 54294 MW; B524D1430B3C9075 CRC64;

Alignment Scores:
Pred. No.: 6,06e-44 Length: 486
Score: 614.50 Matches: 155
Percent Similarity: 48.22% Conservative: 62
Best Local Similarity: 34.44% Mismatches: 163
Query Match: 16.42% Indels: 70
DB: 4 Gaps: 12

US-09-645-078-1 (1-2043) x Q9NS84 (1-486)
QY 7 CATTGTGTGGTACGGAGACGACAGAAAGGTTAGAGAGAAAGCGCGCGGCTA 66
Db 44 HisCysProGlyLeuGlnArg-----SerLeu 52
QY 67 GCAGTG---AGCCTCAAAAGCAGCGAGGAGCCCAAGCCACAAAGGTCTTCCACTTCAG 123
Db 53 GlyValTrpSerLeuGluAlaAlaAlaGlyGluArgGluGlnGlyAlaGluAlaArg 72
QY 124 CACAATGCTACTGCCTAAAAAATGAAGCTCCTGCTGTTCTGCTGTTCCAGATGGCCAT 183
Db 73 AlaAlaGluGlyGlyAlaAsnGlnSerProArg-----PheProSerAsnLeu 89
QY 184 CTGGCTTATCTTCCAGGTACAGTACAGCCACCAATCAGCTCCCTGCTATGAGGCACA 243
Db 90 SerGlyAlaValGlyGluAlaVal-SerArg----- 99
QY 244 GCCCGAGCGCATGACGCTGCTGTTCTGCTCTCTCTGCGGCTCTGCTGCTTTTGTGGG 303
Db 100 ----GluLysGlnHisIleTyrValHisAlaThrTrpArgThrGlySerSerPheLeuGlu 118
QY 304 GCAGCTTTTGGGACGACCCAGATGTTTCTACCTGATGGAGCCCGCTGCGACGTGTG 363
Db 118 ygluLeuPheAenGlnHisProAspValPheTyrLeuTyrGluProMetTrpHisLeuTr 138

QY 364 GATGACCTTCAAGCAGACCGCCTGGATGTCACATGGCTGTGGGATCTGATACG 423
Db 138 pGlnAlaLeuTyrProGlyAspAlaGluSerLeuGlnGlyAlaLeuArgAspMetLeuAr 158
QY 424 GCGCGCTTCTTGTGGACATGAGCCTCTTGTATGCCTACATGGAACCTGGT---CCCGG 480
Db 158 gSerLeuPheArgCysAspPheSerValLeuArgLeuTyrAlaProProGlyAspProAl 178
QY 481 GAGACAG-----TCCAGCCTCTTTCAGTGGGAGAACAG 513
Db 178 aAlaArgAlaProAspThrAlaAsnLeuThrAlaAlaLeuPheArgTrpArgThrAs 198
QY 514 CCGGCGCTGTGTTCTGACCTGCTGTGACATCATCCCAAGATGAA----- 562
Db 198 nLysValIleCysSerProProLeuCysProGlyAlaProArgAlaArgAlaGluValG 218
QY 563 -ATCATCCCCGGGCTCAGTGCAGGCTCTGTCGATCAACAGCCCTTTCAGTGGTGG 621
Db 218 yLeuValGluAspThrAlaCysGluArgSerCysProProValAlaIleArgAlaLeuG 238
QY 622 GAAGCGCTCGCGCTCTACAGCCAGCGGTGCTCAAGGAGTGGCTTCTCAACCTGCA 681
Db 238 uAlaGluCysArgLysTyrProValValIleLysAspValArgLeuLeuAspLeuG 258
QY 682 GTCCCTCTTACCCGCTGTGAAGACCCCTCCCTCACTCATATCTGTCACCTGCTCG 741
Db 258 yValLeuValProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheAr 278
QY 742 GGACCCCGCGCGCTGCTCCGTTCCCGAGAACGACAAAGGAGATCTCATGTTGACAG 801
Db 278 gAspProArgAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgGluSe 298
QY 802 T-----CGATTGTGATG----- 814
Db 298 rIleGlnValLeuArgThrArgGlnArgGlyAspArgPheHisArgValLeuLeuAlaHi 318
QY 815 -----GGGAGCATGAGCAAAACTCAAGAGGAGGAGGACCAACCCCTA 855
Db 318 sGlyValGlyAlaArgProGlyGlnSerArgAlaLeuProAlaAlaProArgAlaAs 338
QY 856 CTAT-----GTGATGCGAGTCTATCGCCAAAGCCAGCTGGAGATCTACAAGAC 903
Db 338 pPhePheLeuThrGlyAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuLeuPh 358
QY 904 CATCCAGCTCTTGGCCCAAGGCCCTGCGAGAACGCTACTGCTTGTGCGCTATGAGGACCT 963
Db 358 eAlaArgGlyAlaProAlaTrpLeuArgArgArgTyrLeuArgLeuArgTyrGluAspLe 378
QY 964 GGCTCGAGCCCTGTGGGCCAGACTTCCCGGAATGTATGAATTCGTGGGATTTGSAATCTT 1023
Db 378 uValArgGlnProArgAlaGlnLeuArgArgLeuLeuArgPheSerGlyLeuArgAlaLe 398
QY 1024 GCGCATCTTACAGCCTGGTGCATACATCACCGAGGCAAGGCGCATGGT---GACCA 1080
Db 398 uAlaAlaLeuAspAlaPheAlaLeuAsnMetThrArgGlyAlaAlaTyrGlyAlaAspAr 418
QY 1081 CGCTTTCACACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTTGGCGCTGCTCTTT 1140
Db 418 gProPheHisLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrpArgGluArgLe 438
QY 1141 GCCCTATGAAAGTTTCTGACITTCAGAAACCTGTGCGCATGCCATGAATTTGCTGGG 1200
Db 438 uSerArgGluGlnValArgGlnValGluValAlaCysAlaProAlaMetArgLeuLeuAl 458
QY 1201 CTACGCCACGTCAGATCTGAACAGAA 1228
Db 458 atyr-----ProArgSerGlyGluGlu 465

RESULT 13
O88276 PRELIMINARY; PRT; 530 AA.
ID O88276
AC O88276
DT 01-NOV-1998 (TrEMBLrel. 08, Created)

RA Jones S.J., Marra M.A.;
 RT "generation and initial analysis of more than 15,000 full-length human
 RI cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2].
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR ENBL; BC051963; AAHS1963.1; -.
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00585; Sulfotransfer; 1.
 KW Transferase.
 SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Alignment Scores:	
Pred. No.:	2,69e-42
Score:	595.50
Percent Similarity:	54.4%
Best Local Similarity:	36.54%
Query Match:	15.91%
DB:	11
	Gaps:
	6
	Indels:
	29
	Mismatches:
	137
	Conservative:
	65
	Matches:
	133
	Length:
	530

US-09-645-078-1 (1-2043) x Q60WV3 (1-530)

248 QY GAGCGCATCGAGTGTGGTCTGCTCTTCCTGGCGCTCTGGCTCTCTTTTGTGGGCGAG 307
163 Db LysArgGlnLeuValTyValPheThrThrTrpArgSerGlySerPhePheGlyGlu 182
308 QY CTTTITGGCGACGACCCAGATGTTTTCACCTGATGGAGCCCGCTCGCACGTGTGGATG 367
183 Db LeuPheAsnGlnAsnProGluValPhePheLeuTyGluProValTrpHisValTrpGln 202
368 QY ACCTTCACGACGACGCGCTGGATGCTGCATGCTGTCGGGATCTGATACCGGCC 427
203 Db LysLeuTyProGlyAspAlaValSerLeuGlnGlyAlaAlaArgAspMetLeuSerAla 222
428 QY GTCTCTCTTGGGACATGAGCGCTCTTTTCATCGCTACATGGAACCTGTCTCCCGGACAGAC 487
223 Db LeuTyArgCysAspLeuSerValPheGlnLeuTySerProAlaGlySerGlyGlyArg 242
488 QY TCC-----AGCCTCTTCAGTGGGAGAACAGCCGGGCCCTGTCTTCGTGCACCT 535
243 Db AsnLeuThrThrLeuGlyLePheGlyAlaAlaThrAsnLysValValCysSerSerPro 262
536 QY GCCTGTGACATCATCCCAAGATGAATCATCCCCGGGCTCACTCAGCGCTCTCTG--- 592
263 Db LeuCysProAlaTyx---ArgLysGluValValGlyLeuValAspAspArgValCysLys 281
593 QY ---TGAGTCAACACCCCTTTGAGGTGTGGAGAGCCCTCGCTCTACAGCCACGCTG 649
282 Db LysCysProGlnArgLeuAlaArgPheGluGluGlyCysArgLysTyArgThrLeu 301
650 QY GTGCTCAAGAGGTGCGTCTTTCACCTGACGTCCCTCTACCGCTGTCTGAAGACCCCT 709
302 Db ValLeuLysGlyValArgValPheAspValAlaValLeuAlaProLeuLeuLysAspPro 321
710 QY TCCCTCAACCTCATATGTCGACCTGTGCGGAGCCCGCGGCGCTGTCCTGTCCTCCGA 769
322 Db AlaLeuAspLeuLysValLeuLeuValArgAspProArgAlaValAlaSerSerArg 341
770 QY GAACGCAAAAGGAGATCTCATGATTCGACAGTCGCATTTGTGATG----- 814
342 Db IleArgSerArgHisGlyLeuIleArgGluSerLeuGlnValValArgSerArgAspPro 361
815 QY -----GGGAGCATGAGCAAAACTCTCAGAGAGAG 844
362 Db ArgAlaHisArgMetProPheLeuGluAlaAlaGlyHisLysLeuGlyAlaLysLysGlu 381
845 QY GACCAA-----CCCTACTAT-----GTGATGCAAGTCTATCTCCCAAGC 883

D	b		382 GlyMetGlyGlyProAlaAspTyrHisAlaLeuGlyAlaMetGluValIleCysAsnSer	401	
Q	y		884 CAGCTGGAGATCTACAAGACCATCCAGTCCTTTGCCCAAGGCCCTGCAGGAACGCTACTCGT	943	
D	b		402 MetalalysThrLeuGlnThrAlaLeuGlnProPrcAspTrpLeuGlnGlyHisTyLeu	421	
Q	y		944 CTGTGGCCCTATGAGGACCCTGGCTGAGGCCCTGTGGCCAGACATTCCCGAATGTTATGAA	1003	
D	b		422 ValValAlaGtyrGluAspLeuValGlyAspProValLysThrLeuArgValTyrAsp	441	
Q	y		1004 TTCGTGGATTTGGAAATCTCTTGCCCATCTTCAGACCTGGGTGCATACATCACCGGAGGC	1063	
D	b		442 PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly	461	
Q	y		1064 AAGGCGATGGTGACCACGGCTTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAG	1123	
D	b		462 SerGlySerSerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAsn	481	
Q	y		1124 GCTTGGCGCTGGCTTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAGACCTGTGGCGAT	1183	
D	b		482 AlaTprArgThrAlaLeuThrPheGlnGlnIleLysGlnValGluGluPheCysTyrGln	501	
Q	y		1184 GCATGAATTTTCGGGCTACCGCCACGTCAGATCTGAACAGAACACAGAAACCTGTGTG	1243	
D	b		502 ProMetalValLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer	521	
Q	y		1244 CTGGATCTCTGTG	1255	
D	b		522 LysThrLeuLeu	525	
R	E	S	RESULT 15		
Q	U	E	D5 Q9UED5 PRELIMINARY; PRT; 483 AA.		
I	C		Q9UED5;		
C			Q9UED5;		
D	T		01-MAY-2000 (TREMBLrel. 13, Created)		
D	T		01-JUN-2003 (TREMBLrel. 23, Last sequence update)		
D	T		01-JAN-2003 (TREMBLrel. 24, Last annotation update)		
D	E	N	-Acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).		
G	N		G6ST.		
O	S		Homo sapiens (Human).		
O	C		Eukaryota; Metazoa; Chordata; Vertebrata; Eureleostomi;		
O	X		Mammalia; Eutheria; Primates; Catarrhini; Homniidae; Homo.		
O	X		NCBI_TaxID=9606;		
R	N	[1]	SEQUENCE FROM N.A.		
R	P		TISSUE=Brain;		
R	C		MEDLINE=98391845; PubMed=9722682;		
R	A		Uchimura K., Muramatsu H., Kanane T., Ogawa H., Yamakawa T., Fan Q.,		
R	A		Mitsuka K., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,		
R	A		Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,		
R	T		*Human N-acetylglucosamine-6-O-sulfotransferase involved in the		
R	T		biosynthesis of 8-sulfo s-lally Lewis x: Molecular cloning, chromosomal		
R	T		mapping, and expression in various organs and tumor cells.";		
R	J		J. Biochem. 124:670-678(1998).		
D	R		EMBL; AB014679; BAA34265.2 ;		
D	R		GO; GO:0008146; F:sulfotransferase activity; IEA.		
D	R		GO; GO:0016740; F:transferase activity; IEA.		
D	R		InterPro: IPR000863; Sulfotransferase.		
D	R		Pfam: PF00565; Sulfotransfer; 1.		
K	M		Transferase.		
S	Q		SEQUENCE 483 AA; 52787 MW; B27797D4931BA18 CRC64;		
A			Alignment Scores:		
P	r	e	d. No.:	Length: 483	
S	c	o	r	e:	Matches: 132
B	e	s	t Local Similarity:	Conservative: 66	
Q	u	e	ry Match:	Mismatches: 137	
D	b	:		Indels: 29	
				Gaps: 6	
U	S	-09-645-078-1 (1-2043) x Q9UED5 (1-483)			
Q	y		248 GAGCGCATGCGACGTGCTGGTTCGTCTTCTCTCGCGCTCGCTGCTCTCTTTGTGGGGCAG	307	

QY 248 GAGCGCATGCACGTGCTGGTTCTGTCTTCCCTGGCGCTCTGGCTCTTCTTTTGTGGGCAG 307

```

Db      116 LysArgGlnLeuValTyrValPheThrThrTrpArgSerGlySerPhePheGlyGlu 135
QY      308 CTTTGGGAGACACCAAGATGTTTCTACCTGATGGAGCCCGCTGGCACTGTGGATG 367
Db      136 LeuPheAsnGlnAsnProGluValPhePheLeuTyrGluProValTrpHisValTrpGln 155
QY      368 ACCTTCAAGCAGACGACCCCTGGATGCTGCATCGCTGTGGGATCTGTATCGGGCC 427
Db      156 LysLeuTyrProGlyAspAlaValSerLeuGlnGlyAlaAlaArgAspMetLeuSerAla 175
QY      428 GTCTTCTTGTGCACATGAGCGTCTTTGATGCTCATGGAACCTGTCTCCCGGAGACAG 487
Db      176 LeuTyrArgCysAspLeuSerValPheGlnLeuTyrSerProAlaGlySerGlyGlyArg 195
QY      488 TCC-----AGCTCTTTTCAGTGGGAGAACAGCCGGCCCTGTGTTCTGCACCT 535
Db      196 AsnLeuThrThrLeuGlyIlePheGlyAlaAlaThrAsnLysValValCysSerSerPro 215
QY      536 GCCTGTGACATCATCCCAAGATGAATCATCCCCCGGCTCTACCTGACGGCTCCTG--- 592
Db      216 LeuCysProAlaTyr--ArgLysGluValValGlyLeuValAspArgValCysLys 234
QY      593 ---TGACGTCAACAGCCCTTGAGTGTGGAGAGCCCTGCGCTCTACAGCCACGCTG 649
Db      235 LysCysProProGlnArgLeuAlaArgPheGluGluGluCysArgLysTyrArgThrLeu 254
QY      650 GTCTCAAGGAGGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAAGACCCC 709
Db      255 ValIleLysGlyValArgValPheAspValAlaValLeuAlaProLeuLeuArgAspPro 274
QY      710 TCCTCTAACCTGCATATCGTGCACCTGTGTCGGGACCCCGGCGCTGTTCGTTCCCGA 769
Db      275 AlaLeuAspLeuLysValIleHisLeuValArgAspProArgAlaValAlaSerSerArg 294
QY      770 GAACGCACAAAGGAGATCTCATGATTGACAGTCGCATTGTGATG----- 814
Db      295 IleArgSerArgHisGlyLeuIleArgGluSerLeuGlnValValArgSerArgAspPro 314
QY      815 -----GGCAGCATGAGCAAAACTCAAGAAGGAG 844
Db      315 ArgAlaHisArgMetProPheLeuGluAlaAlaGlyHisLysLeuGlyAlaLysLysGlu 334
QY      845 -----GACCAACCTACTAT-----GTGATGAGGTCTATCTGCCAAAGC 883
Db      335 GlyValGlyGlyProAlaAspTyrHisAlaLeuGlyAlaMetGluValIleCysAsnSer 354
QY      884 CAGCTGGAGATCTACAGACCATCCAGTCTTCCCAAGGCCCTGCGAGAACGCTACCTG 943
Db      355 MetAlaLysThrLeuGlnThrAlaLeuGlnProProAspTrpLeuGlnGlyHisTyrLeu 374
QY      944 CTTGTGGCTATGAGGACCTGGCTCGAGCCCTGTGCCCCAGACTTCCCGAATGTATGAA 1003
Db      375 ValValArgTyrGluAspLeuValGlyAspProValLysThrLeuArgArgValTyrAsp 394
QY      1004 TTCGTGGGATTGGAATTCCTTCCCTCATCTTCAGACCTGGGTGCATACATCACCCGAGGC 1063
Db      395 PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly 414
QY      1064 AAGGCGATGGTGACCGCTTCCACAAATCCAGGATGCCCTTAATGTCTCCCGAG 1123
Db      415 SerGlySerSerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn 434
QY      1124 GCTTGGCGCTGTCTTTCCTCTATGAAAGGTTTCTCGACTTCAGAAAGCTGTGGCGAT 1183
Db      435 AlaTrpArgThrAlaLeuThrPheGlnGlnIleLysGlnValGluGluPheCysTyrGln 454
QY      1184 GCCATGAATTGCTGGGCTACCGCCACGTGAGATCTGAACAGAACAGAAACCTGTTG 1243
Db      455 ProMetAlaLeuLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer 474
QY      1244 CTGGATCTTCTG 1255

```

Db 475 LysThrLeuLeu 478

Search completed: May 6, 2004, 11:05:46
Job time : 124 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: May 6, 2004, 10:52:03 ; Search time 96.5 Seconds
(without alignments)
11963.618 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 3742
Sequence: 1 gaattccattgttgggta.....tgggattccataaaaaaaaaa 2043

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool -P/US09645078/runat 06052004 104558 21717/app query.fasta_1.2183
-DB=A Geneseq 29Jan04 -QMT=fastan -SUFFIX=xag -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR VAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09645078 @CGN 1.1_B1 @runat 06052004 104558 21717 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq 29Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004as: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	54.5	386	2	AAY39918 Human gly
2	2035	54.4	386	2	AAM93309 Human pol
3	2008	53.7	380	5	Aau11274 Human L-s
4	1984	53.0	386	3	Aay79219 Human tra
5	1519.5	40.6	388	2	Aay39919 Mouse gly
6	1208.5	32.3	418	3	Aab41947 Human ORF
7	1030.5	27.5	418	5	ABD81557 Mouse int
8	1028.5	27.5	395	4	Aay72640 Human gly
9	1028.5	27.5	395	5	ABD81554 Human cor
10	1028.5	27.5	395	5	AAE15438 Human dru

11	1020	27.3	395	4	AAY72638	Aay72638 Mouse gly
12	1020	27.3	395	5	AAU11275	Aau11275 Murine in
13	1017.5	27.2	395	5	ABB81555	Abb81555 Consensus
14	1009	27.0	390	5	AAY72639	Aay72639 Human gly
15	1009	27.0	390	5	ABB81556	Abb81556 Human int
16	821	16.4	486	5	ABB81560	Abb81560 Human hig
17	614.5	16.4	486	5	ABP56121	Abp56121 Human cho
18	598.5	16.0	483	2	RAY31656	Ray31656 Mouse N-a
19	591.5	15.8	530	4	AA895367	Aa895367 Human pro
20	587.5	15.7	484	2	RAY31657	Aay31657 Human N-a
21	587.5	15.7	531	5	AAU69414	Aau69414 Lung smal
22	549	14.7	411	2	AAW61100	Aaw61100 Keratan s
23	549	14.7	411	2	AAE25356	Aae25356 Human cho
24	549	14.7	411	6	ABU03503	Abu03503 Angiogene
25	519.5	13.9	169	5	ABB81558	Abb81558 Human cor
26	511.5	13.7	169	5	ABB81559	Abb81559 Human int
27	500.5	13.4	458	2	AAW06480	Aaw06480 Chick cho
28	492	13.1	499	6	ABR41139	Ab41139 Human DIT
29	482	12.9	479	2	AAW52863	Aaw52863 Glycosami
30	336	9.0	169	5	ABB81561	Abb81561 Human N-a
31	335.5	9.0	179	5	ABB81562	Abb81562 Human ker
32	293	7.8	174	5	ABB81563	Abb81563 Human cho
33	215	5.7	363	4	ABB64512	Abb64512 Drosophil
34	205	5.5	1212	2	AAW87503	Aaw87503 Human N-m
35	199.5	5.3	1061	2	AAW87504	Aaw87504 Human N-m
36	198.5	5.3	1938	6	ABB98398	Abb98398 Streptomy
37	198	5.3	1232	3	AAE28239	Aae28239 Human N-m
38	198	5.3	1232	5	AAW47961	Aaw47961 Human NMD
39	198	5.3	1232	6	ABU61439	Abu61439 Human N-m
40	198	5.3	1232	7	AAE39250	Aae39250 Human NMD
41	192.5	5.1	1081	3	AAE26240	Aae26240 Human N-m
42	192.5	5.1	1081	5	ABB56523	Abb56523 Human NMD
43	192.5	5.1	1081	6	ABU61440	Abu61440 Human N-m
44	192.5	5.1	1081	6	AAE39251	Aae39251 Human NMD
45	187.5	5.1	19938	6	ABB98398	Abb98398 Streptomy

ALIGNMENTS

RESULT 1
AAY39918
ID AAY39918 standard; protein; 386 AA.
XX
AC AAY39918;
XX
XX 08-DEC-1999 (first entry)
XX Human glycosyl sulfotransferase-3 protein sequence.
DE Human glycosyl sulfotransferase-3 protein sequence.
XX
XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; human;
KW secondary lymph organ.
XX
OS Homo sapiens.
XX
XX WO9949018-A1.
XX
XX 30-SEP-1999.
XX
XX 26-FEB-1999; 99WC-US004316.
XX
XX 20-MAR-1998; 98US-00045284.
XX
XX 12-NOV-1998; 98US-00190911.
XX
XX (REGC) UNIV CALIFORNIA.
XX (SYNT) SYNTEX USA INC.
XX
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX WPI; 1999-580442/49.
XX N-PSDB; AAZ20792.
XX
XX Human and murine glycosyl sulfotransferase 3 and related polynucleotides.

FT	Modified-site	308	/note= "potential N-glycosylation"
FT	Modified-site	329	/note= "potential N-glycosylation"
FT	Modified-site	364	/note= "potential O-phosphorylation"
FT	Modified-site	380	/note= "potential O-phosphorylation"
XX	WC200014251-A2.		
XX	16-MAR-2000.		
XX	09-SEP-1999;	99WO-US020989.	
XX	10-SEP-1998;	98US-00150657.	
PR	04-NOV-1998;	98US-001866779.	
PR	11-MAY-1999;	99US-0133642P.	
XX	(INCY-) INCYTE PHARM INC.		
XX	Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;		
PI	Hillman JL, Azimzai Y;		
XX	WPI: 2000-256996/22.		
DR	N-PSDB; AAZ94211.		
XX	Human transferase proteins useful for preventing, diagnosing and treating cancers and developmental, gastrointestinal, genetic, immunological, neurological, reproductive and smooth muscle disorders.		
XX	Claim 1; Page 90-91; 113pp; English.		
XX	The present sequence is that of human transferase TRNSFS-11, 1 of 15 claimed novel human transferase proteins of the invention (see AAY79203-23). The sequence was deduced from a cDNA clone (see AAZ94211) isolated from a galbladder library. It shows homology to mouse N-acetylglycosamine 6-O-sulfotransferase. TRNSFS-11 is expressed in dermatologic and gastrointestinal tissues, especially those associated with inflammation and cell proliferation. The new human transferases and polynucleotides can be used in the diagnosis, prevention and treatment of cancer, developmental disorders, gastrointestinal disorders, genetic disorders, immunological disorders, neurological disorders, reproductive disorders, and smooth muscle disorders. The polypeptides can also be used to raise antibodies, and to screen for agonists and antagonists of transferase activity		
XX	Sequence 386 AA;		
XX	Alignment Scores:		
XX	Pred. No.:	6,84e-168	Length: 386
XX	Score:	1984.00	Matches: 383
XX	Percent Similarity:	99.48%	Conservative: 2
XX	Best Local Similarity:	98.97%	Mismatches: 1
XX	Query Match:	53.02%	Indels: 2
XX	DB:	3	Gaps: 0
XX	US-09-645-078-1 (1-2043) x AAY79219 (1-386)		
QY	128	ATGCTACTGCCTAAAAAATGAAGCTCTGCTGTTCTGTTCCAGATGGCCATCTTG 187	
Db	1	MetLeuLeuProLysLysMetLysLeuLeuLeuPheLeuValSerGlnMetAlaLeu 20	
QY	188	GCTATTCTTCCACATGACAGGCACACATCAGCTCCCTGTGTATGAAGGCACAGCCC 247	
Db	21	AlaLeuPhePheHisMetTy-SerHisAsnIleSerLeuSerMetLysAlaGlnPro 40	
QY	248	GAGCCATGCAGTCTGGTCTGCTCTTCCTGGCGCTCTGGCTCTCTTTGTTGGGGCAG 307	
Db	41	GluArgMetHisValLeuValLeuSerIrpArgSerGlySerSerPheValGlyGln 60	
QY	308	CTTTTGGGCACACCCAGATGTTTCTACCTGATGGAGCCCGCTGGACGTTGTGGATG 367	

XX Mouse glycosyl sulfotransferase-3 protein sequence.

DE Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.
XX
OS Mus sp.
XX
PN WO9949018-A1.
XX
PD 30-SEP-1999.
XX
XX 26-FEB-1999; 99WO-US004316.
XX
XX 20-MAR-1998; 98US-00045284.
PR 12-NOV-1998; 98US-00190911.
XX
XX (REGC) UNIV CALIFORNIA.
PA (SYNT) SYNTEX USA INC.
XX
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX
DR WPI; 1999-580442/49.
DR N-PSDB; AAZ20793.
XX
PT Human and murine glycosyl sulfotransferase 3 and related polynucleotides.
XX
XX Claim 2; Fig 4; 59pp; English.
XX
CC This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of the
CC invention. The nucleic acid sequences, probes and primers derived from
CC these proteins and antibodies are useful in detecting homologues. The
CC sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs
XX
SQ Sequence 388 AA;

Alignment Scores:

Pred. No.:	1 98e-126	Length:	388
Score:	1519.50	Matches:	283
Percent Similarity:	85.13%	Conservative:	49
Best Local Similarity:	72.56%	Mismatches:	55
Query Match:	40.61%	Indels:	3
DB:	2	Gaps:	2

US-09-645-078-1 (1-2043) x AA393919 (1-388)

Qy 128 ATGCTACTGCCTAAAAAATGAAGTCTCGTGTTCTGGTTTCACAGATGCCATCTTG 187

Dd 1 MetMetLeuLysLysGlyArgLeuLeuMetPheLeuGlySerGlnValIleValVal 20

Qy 188 GCTCTATTCTTCACATGTACAGCCACACATCATGCTCCCTGTCTATGAAGGCACAGCCC 247

Dd 21 AlaLeuPheIleHisMetSerValHisArg-----HisLeuSerGlnArgGluGluSer 38

Qy 248 GAGCGC---ATGCACGTGCTGGTCTGCTTCCTGGCGCTCGCTCTCTTTTGTGGG 304

Dd 39 ArgArgProValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGly 58

Qy 305 CAGCTTTTTGGGAGCACCCAGATGTTTTCTACCTGATGGAGCCCGCTGGCAGCTGTGG 364

Dd 59 GlnLeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrp 78

Qy 365 ATGACCTTCAAGCAGACCGCTGGATGCTGCACATGGCTGTGCGGATCTGATACGG 424

Dd 79 MetThrPheThrSerSerThrAlaTrpLysLeuHisMetAlaValArgSpleuLeuArg 98

Qy 425 GCCTGCTTCTTGTCGACATGAGCGTCTTTGATGCTCATGGAACTGGTCCCCGGAGA 484

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; Gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX Homo sapiens.
 OS
 XX
 XX WO200058473-A2.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 31-MAR-2000; 2000WO-US008621.
 PF
 XX
 XX 31-MAR-1999; 99US-0127607P.
 PR
 PR 02-APR-1999; 99US-0127636P.
 PR
 PR 05-APR-1999; 99US-0127728P.
 PR
 PR 30-MAR-2000; 2000US-00540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX
 XX Shinkets RA, Leach M;
 PI
 XX
 XX WPI; 2000-602362/57.
 DR
 DR N-PSDB; AAC76156.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 XX Claim 11; Page 2599-2600; 5507pp; English.

CC AAC74446 to AAC7506 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
 CC anticongulant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antinaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 418 AA;

Alignment Scores:
 Pred. No.: 1,17e-98 Length: 418
 Score: 1208.50 Matches: 249
 Percent Similarity: 69.34% Conservative: 54
 Best Local Similarity: 56.98% Mismatches: 85
 Query Match: 32.30% Indels: 49
 DB: 3 Gaps: 10

US-09-645-078-1 (1-2043) x ABA41947 (1-418)

QY 29 GACAGAGGGTAGAGAGAAAGCGCA-----TGGCCGGGCTAGCAGTGAGCCTCT 79
 |||:|||||
 DB 10 ASPLAArgCysSerSerProSerSerProGlyArgTirPro----- 23

QY 80 CAAAGACAGCAGGGAAGGCCCAAGCCCAAGGCTTCT-----CACTTCAGCACA----- 127
 |||:|||||
 DB 24 -----ProAlaAlaArgMetTirLeuProArgPheSerSerIysThr 37
 QY 128 -----ATCCTACTGCTTAAAAAATGAAGCTCTCTGCTGTTCTGTTTCCAGATG 178
 |||:|||||
 DB 38 ValThrValLeuLeuAlaGlnThrThrCysLeuLeuLeuPheIleIleSerArgPro 57
 QY 179 GCATCTTGGCTCTATTCTTCCACATGTACAGCCACAAATCATCAGCTCCCTGTCTTGAAG 238
 |||
 DB 58 -----GlyProSerSerProAla 63
 QY 239 GCACAGCCGAGGCGATGACGTGTGTGTCTTCTCTGCGGCTCTGGCTTCTTCTTT 298
 |||:|||||
 DB 64 GlyGlyGluAspArgValHisValLeuValLeuSerSerTirArgSerGlySerSerPhe 83
 QY 299 GTGGGCGAGCTTTTGGCGAGCACCAGATGTTTCTTACCTCATGGAGCCCGCTGGCAC 358
 |||:|||||
 DB 84 LeuGlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTirHis 103
 QY 359 GTGTGGATGACCTTCAAGCAGAGCAGCCGCTGGATGTGCATGCTGTGGGATCTG 418
 |||:|||||
 DB 104 ValTirThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeu 123
 QY 419 ATACGGGCGCTCTTGTGTGCGACATGAGCGCTTTGTGTGTACATGAGAACTGTCTCC 478
 |||:|||||
 DB 124 MetArgSerIlePheLeuCysAspMetAspValPheAspAlaTyrMetGluProGlyPro 143
 QY 479 CGGAGACAGCTCCAGCCTCTTTCAGTGGGAGAACAGCCGGCGCTGTGTCTGCACCTGCC 538
 |||:|||||
 DB 144 ArgArgGlnSerSerLeuPheGlnTirGluAsnSerArgAlaLeuCysSerAlaProAla 163
 QY 539 TGTGACATCATCCACAGATGAATCATCCCGGGCTCACTGAGGCTCTGTGTGCACT 598
 |||:|||||
 DB 164 CysAspIleIleProGlnAspGluIleIleProAlaHisCysArgLeuLeuCysSer 183
 QY 599 CAACAGCCCTTTGAGGTGTGGAGAGGCTCCCGCTCTTACAGGCACAGCTGTGTCTCAAG 658
 |||:|||||
 DB 184 GlnGlnProPheGluValValGluLysAlaCysArgSerTyrSerHisValValLeuLys 203
 QY 659 GAGTGGCTTCTTAACTGTGAGTCTCTTACCCTGCTCTGAAGACCCCTCCCTCAAC 718
 |||:|||||
 DB 204 GluValArgPhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsn 223
 QY 719 CTGCATATCTGCACCTGTGTCCGGGACCCCGGCGCTGTCTCCGTTCGAGAAAGCACA 778
 |||:|||||
 DB 224 LeuHisIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluAlaAla 243
 QY 779 AAGGAGATCTCATGATTGACATGCGATGTGTGATGGGCGCAGCATGAGCAAAACTCAAG 838
 |||:|||||
 DB 244 GlyProIleLeuAlaArgAspAsnGlyIleValLeuGlyThrAsn---GlyLysTirVal 262
 QY 839 AAGGAGGACCAACCTTACTATGTGTGAGGTGATCTGCAAGGCGCAGCTGGAGATCTAC 898
 |||:|||||
 DB 263 GluAlaAspProHisLeuArgLeuIleArgGluValCysArgSerHisValArgIleAla 282
 QY 899 AAG-----ACCATCCAGTCTTTCGAGGAGGCGCTGAGGAAAGCTGTGTGTGGCGC 952
 |||:|||||
 DB 283 GluAlaAlaThrLeuLysPro---ProProPheLeuArgGlyArgTyrArgLeuValArg 301
 QY 953 TATGAGGACCTGCTGAGCCCTGTGCGCCAGACTTCCGGAATGTATGAATTCGTGGGA 1012
 |||:|||||
 DB 302 PheGluAspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThrGly 321
 QY 1013 TTGGAATTCCTGCGCCCATCTTTCAGACTGGGTGTCATTAACATCACCCGAGGAGGCGCATG 1072
 |||:|||||
 DB 322 LeuThrLeuThrProGlnLeuGluAlaTirIleHisAsnIleThrHisGlySerGlyIle 341
 QY 1073 GGTGAC-----CACGCTTTCCACAAATGCCAGGATGCCCTTATGCTCTCCAGGCT 1126
 |||:|||||
 DB 342 GlyLysProIleGluAlaPheHisThrSerSerArgAsnAlaArgAsnValSerGlnAla 361
 QY 1127 TGGCGCTGGTCTTTGGCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGCGCATGCC 1186

Db 362 TrpArgHisAlaLeuProPheThrLysileLeuArgValGlnGluValCysAlaGlyAla 381
 QY 1187 ATGAATTGGCTGCTACCGCCAGTCAGATCTGAACAAGACAGAGAAACCTGTTGCTG 1246
 Db 382 LeuGlnLeuLeuGlyTyrArgProValTyrSerAlaaspGlnGlnArgaspLeuThrLeu 401
 QY 1247 GATCTTCTG-----TCTACCTGGACTGTCCTCGAG 1276
 Db 402 AspLeuValLeuProArgGlyProAspHisPheSerTrpAlaSerProAsp 418

RESULT 7

ABB81557
 ID ABB81557 standard; protein; 418 AA.

AC ABB81557;
 DT 05-SEP-2002 (first entry)

DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.

KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.

OS Mus musculus.

XX US2002061562-A1.

PN 23-MAY-2002.

XX 09-AUG-2001; 2001US-00927602.

XX 11-AUG-2000; 2000US-00638211.

PR 11-AUG-2000; 2000US-0325773P.

XX (FUKU/) FUKUDA M N.

PA (AKAMA/) AKAMA T O.

PI Fukuda MN, Akama TO;

XX WPI; 2002-507643/54.

DR New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.

XX Example 5; Page 24-25; 69pp; English.

CC The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratoplasty or keratectomy. The present sequence
 CC represents mouse intestinal N-acetylglucosamine-6- sulfotransferase,
 CC which is given in comparison with (I) in the exemplification of the
 CC present invention

XX Sequence 418 AA;

Alignment Scores:

Pred. No.: 9,07e-83 Length: 418
 Score: 1030.50 Matches: 221
 Percent Similarity: 66.75% Conservative: 44
 Best Local Similarity: 55.67% Mismatches: 120
 Query Match: 27.54% Indels: 12
 DB: 5 Gaps: 7

US-09-645-078-1 (1-2043) x ABB81557 (1-418)

QY 128 ATGCTACTGCCTAA-----AAATGAAGCTCCTGCTGTTCTTGGTTCCAGATGCC 181
 Db 25 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuMetValGlnThrGly 44
 QY 182 ATCTTGGCTCTATTCTTCCACATGTACAGCCACACATCAGTCCTCTCTATGAAGSCA 241
 Db 45 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 60
 QY 242 CAGCCCGAGCGCATCAGCTGCTGTTCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 301
 Db 61 LeuGlyGluArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheVal 80
 QY 302 GGGCAGCTTTTGGCAGCACCAGATGTTTCTACTGTATGGAGCCCGCTGGCACCTG 361
 Db 81 GlyGlnLeuPheSerGlnHisProaspValPheTyrLeuMetGluProAlaTrpHisVal 100
 QY 362 TGGATGACCTTCAAGCAGACACCGCTGGATGTGCATCGCTGCTGGGATCTGATA 421
 Db 101 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeuLe 120
 QY 422 CGGGCGCTCTTCTGCGACATGAGCGTCTTGTATGCTTACATGGAACCTGGTCCCGG 481
 Db 121 ArgSerValPheLeuCysaspMetaspValPheAspAlaTyrLeu---ProTrpArg 139
 QY 482 AGACACTCCAGCTCTTTTCAGTGGGAGAACACCGCGGCTGCTGTCTGCTGACCTGCTGT 541
 Db 140 AsnIleSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProValCys 159
 QY 542 GACATCATCCCAAGATGAATCATCCCGGGCTCACTGCGAGTCTCTGTGCGAGTCAA 601
 Db 160 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLeuPheCysAlaThr 179
 QY 602 CAGCCCTTTGAGGTGTGAGAGAGCGCTGCCCTCTCTACAGCCAGCTGGTCTCAAGGAG 661
 Db 180 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 199
 QY 662 GTGCGCTTCTTCAACCTGAGTCCCTTACCGCTGCTGAAAGACCCCTCCCTCAACCTG 721
 Db 200 ValArgShePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 219
 QY 722 CATATCGTCACCTGCTCCGGGACCCCGGGCGGTGTTCCGTTCCCGAGAACGCAACAAG 781
 Db 220 ArgIleValHisLeuValArgaspProArgAlaValLeuArgSerArgGlnGlnThrAla 239
 QY 782 GGAATCTCATGATTGACAGTCCATGTTGATGGGCGACATGAGCAAAAATCAAGAGAG 841
 Db 240 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---Glu 258
 QY 842 GAGGACCAACCTACTATGTGTGTCAGGTTCATCTGCAAGCCAGCTGGAGATCTACAAG 901
 Db 259 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 278
 QY 902 ACCATCCAGTCTTGGCCAAAGCCCTGCAGGAACGCTACTCTGCTTGTGGCTATGAGGAC 961
 Db 279 AlaLeuHisLysProProPheLeuGlnAspArgTyrArgLeuValArgTyrGluAsp 298
 QY 962 CTGCTCGAGCCCTGCGCCAGACTCCCGAATGTATGAATTCGTGGATGGAAATTC 1021
 Db 299 LeuAlaArgaspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGlyLeu 318
 QY 1022 TTGCCCATCTTTCAGACTCGGTGTCATAACATCACCGGAGGCAAGGGCATGGGT----- 1075
 Db 319 ThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAlaArg 338
 QY 1076 GACCACGTTTCCACAAAATGCCAGGATGCCCTTAATGTCCTCCAGGGCTCGCGCTGG 1135
 Db 339 ArgGluAlaPheLysThrThrSerArgaspAlaLeuSerValSerGlnAlaTrpArgHis 358
 QY 1136 TCTTTGGCTATGAAAGGTTTCTCGACTTCAGAAAGCCCTGTGGCGATGCGCATGTTG 1195
 Db 359 ThrLeuProPheAlaLysIleArgArgValGlnGlnLeuCysGlyGlyAlaLeuGlnLeu 378
 QY 1196 CTGGGCTACCGCCAGCTCAGTCTCTGAACAGACAGAGAACCTGTTGCTGGATCTTCTG 1255

Db 315 ArgArgGluAlaPheLysThrSerArgAsnAlaLeuAsnValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTGGCCATGAAAGGTTCTCGACTTCAGAAAGCGCTGGCGATGCCATGAAT 1192
Db 335 HisAlaLeuProPheAlaLysIleArgArgValGlnGluLeuCyAlaGlyAlaLeuGln 354
QY 1193 TTGCTGGGCTACCGCAGCTCAGACTGAACAAGAACAGAGAAACCTGCTGATCTT 1252
Db 355 LeuLeuGlyTyArgProValTySerGluAspGluGlnArgAsnLeuAlaLeuAspLeu 374
QY 1253 CTGCTACCTGAGCTGCTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGTTTGTCTGCC 1312
Db 375 Val-----LeuPro-----ArgGlyLeuAsnGlyPhe----- 383
QY 1313 ACCTGTGTGCAGCTCAGTCACTTCTCTGAATGCTTCTGAGCCT 1357
Db 384 ThrTrpAla-----SerSerThrAlaSerHisPro 393

RESULT 9
ABB81554
ID ABB81554 standard; protein; 395 AA.
AC ABB81554;
XX
DT 05-SEP-2002 (first entry)
XX Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22.
XX
OS Homo sapiens.
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
XX (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
XX
XX N-PSDB; ABN89506.
XX
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX
XX Claim 13; Fig 1A-D; 69pp; English.
XX
XX The present sequence represents human corneal N-acetylglucosamine-6-
XX sulfotransferase (I), which is able to catalyze sulfation of keratan
XX sulfate (KS). Also described is a method for monitoring the effect of
XX treatments for macular corneal dystrophy (MCD), and detecting
XX susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX ophthalmological activity. (I) can be used to treat or prevent macular
XX corneal dystrophy types I or II. (I) makes possible treatment of MCD
XX without requiring keratinoplasty or keratectomy
XX
XX Sequence 395 AA;

Alignment Scores:
Pred. No.: 1.34e-82 Length: 395
Score: 1028.50 Matches: 224
Percent Similarity: 65.06% Conservative: 46

Best Local Similarity: 53.98% Mismatches: 108
Query Match: 27.49% Indels: 37
DB: 5 Gaps: 10
US-09-645-078-1 (1-2043) x ABB81554 (1-395)
QY 125 ACAATCTACTGCTTAAATAAAGCTCTGCTGTTTCTGTTTCCAGATGCCCATC 184
Db 12 ThrAlaLeuLeuLeuAlaGlnThrPheLeuLeuLeuPheLeuValSerArgPro----- 29
QY 185 TTGGCTCTATTCTTCCACATGTACAGCCACACATCAGCTCCCTCTCTATGAGGCACAG 244
Db 30 -----GlyProSerSerProAlaGlyGly 37
QY 245 CCGAGCGCATGACAGCTGCTGTTCTTCTTCTTCTGCGCTCTGGCTCTTCTTTTGGGG 304
Db 38 GluAlaArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGly 57
QY 305 CAGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAGCCGCTGGCACGTGG 364
Db 58 GlnLeuPheAsnGlnHisProAspValPheTyLeuMetGluProAlaTrpHisValTrp 77
QY 365 ATGACCTTCAAGCAGAGCAGCCGCTGGATGCTGCGACATGGCTGGGGATCTGATAGG 424
Db 78 ThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuValArg 97
QY 425 GCGCTTCTTGTGCGACATGAGCGTCTTTGTGCTATCATGGAACCTGGTCCCGGAGA 484
Db 98 SerValPheLeuCyAspMetAspValPheAspAlaTyLeu---ProTrpArgArgAsn 116
QY 485 CAGTCCAGCCTTTTTCAGTGGGAGACAGCCGGGCGCTGTGTCTGCACTGCTGTCAC 544
Db 117 LeuSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCySerSerProProAlaCySer 136
QY 545 ATCATCCCAAGATGAATCATCCCGGGCTCACTGCGAGCTCCTGTGAGTCAACAG 604
Db 137 AlaPheProArgGlyAlaIleSerSerGluAlaValCysLysProLeuCyAlaArgGln 156
QY 605 CCCTTTGAGGTGGTGGAGAGCCCTGCGCTCTACAGCAGCTGGTCTCAAGAGAGTG 664
Db 157 SerPheThrLeuAlaArgGluAlaCysArgSerTySerHisValValLeuLysGluVal 176
QY 665 CGCTTCTTCAAGCTGAGTCCCTCTACCGCTGTCTGGAAGACCCCTCCCTCAACCTGCAT 724
Db 177 ArgPhePheAsnLeuGlnValLeuTyProLeuLeuSerAspProAlaLeuAsnLeuArg 196
QY 725 ATGCTGCACCTGGTCCGGAGACCCCGGCGGTGTTCCGTTCGAGAGACCGACAAAGGA 784
Db 197 IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys 216
QY 785 GATCTCATGATTGACAGTCGCTTGTGATGGGGCAGCATGAGCAAAATCAAGAAGGAG 844
Db 217 AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla 235
QY 845 GACCAACCTACTATGTATGATGAGTCTATCTGCCAAAGCCAGCTGGAGATCTACAG--- 901
Db 236 AspProGlyLeuArgValValArgGluValCysArgSerHisValArgIleAlaGluAla 255
QY 902 ---ACCATCCAGTCTTGGCCCAAGCCCTGCGAGGAACGCTACTCTGTGCGCTATGAG 958
Db 256 AlaThrLeuLysPro---ProProPheLeuArgGlyArgTyArgLeuValArgPheGlu 274
QY 959 GACCTGGCTCGAGCCCTGCTGCGCCAGACTTCCCGAATGTATGAATTCTGGGATTTGAA 1018
Db 275 AspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyAlaPheThrGlyLeuSer 294
QY 1019 TTCTTGGCCCATCTTCAGACCTGGGTGCTATACATCACCAGGAGCAAGGCGATGGT--- 1075
Db 295 LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCAGCTTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAGCTTGGGCGC 1132
Db 315 ArgArgGluAlaPheLysThrSerSerArgAsnAlaLeuAsnValSerGlnAlaTrpArg 334

QY 545 ATCATCCCAAGATGAATCATCCCGGGCTCACTGCGAGCTCCCTGTCAGTCAACAG 604
DB |||||
QY 137 AlapheProArgGlyAlaIleSerSerGluAlaValCysIysProLeuCysAlaArgGln 156
DB |||||
QY 605 CCCTTTGAGGTGGAGAGAGCCCTGCGCTCTACAGCCACAGTGGTCTCAAGAGGTG 664
DB |||||
QY 157 SerPheThrLeuAlaArgGluAlaCysArgSerTyrSerHisValValLeuLysGluVal 176
DB |||||
QY 665 GCCTTTCTTCAACCTGACGCTCCCTTACCCGCTGTGAAAGACCCCTCCCTCAACCTGCAT 724
DB |||||
QY 177 ArgPhePheAnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeuArg 196
DB |||||
QY 725 ATCTGTGACCTGTGTCGGGAGACCCCGGCGCTGTTCCTCCGAGAACCCAGCAAGAGGA 784
DB |||||
QY 197 IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys 216
DB |||||
QY 785 GATCTCATGATTACAGTCGCGATTGTGATGGGCGAGCATGACCAAACTCAGAAAGGAG 844
DB |||||
QY 217 AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla 235
DB |||||
QY 845 GACCAACCTCTACTATGTATGATGAGGTTCATCTGCCAAGCCAGCTGGAGATCTCAAG--- 901
DB |||||
QY 236 AspProGlyLeuArgValValArgGluValCysArgSerHisValArgIleAlaGluAla 255
DB |||||
QY 902 ---ACCATCCAGTCTTCCCAAGCCCTGCGAGAACGCTGCTGTGCGCTATGAG 958
DB |||||
QY 256 AlaThrLeuLysPro---ProProPheLeuArgGlyArgIleArgLeuValArgPheGlu 274
DB |||||
QY 959 GACCTGGTCGAGCCCTGTGCGCCAGACTTCCCGAATGTATGATTCGTGGGATTGGAA 1018
DB |||||
QY 275 AspleuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThrGlyLeuSer 294
DB |||||
QY 1019 TTCCTGCCCCATCTCAGACCTGGGTGATACATACATCCCGAGGCAAGGCGATGGT--- 1075
DB |||||
QY 295 LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyVala 314
DB |||||
QY 1076 ---GACCACGCTTCCACACAATGCCAGGATGCTTAAATGTCCTCCAGGCTTGGCGC 1132
DB |||||
QY 315 ArgArgGluAlaPheLysThrSerSerArgAsnAlaLeuAsnValSerGlnAlaTrpArg 334
DB |||||
QY 1133 TGGCTTTTCCCTATGAAGAAGTTTCTCGACTTCAGAAAGCTTGGCGATGCCATGAAT 1192
DB |||||
QY 335 HisAlaLeuProPheAlaLysIleArgArgValGlnGluLeuCysAlaGlyValaLeuGln 354
DB |||||
QY 1193 TTGCTGGGCTACCGGACGTCAGATCTGAAACAAGAACACCTGCTGCTGATCTT 1252
DB |||||
QY 355 LeuLeuGlyTyrArgProValTyrSerGluAspGluGlnArgAsnLeuAlaLeuAspLeu 374
DB |||||
QY 1253 CTGCTACTGAGTGTCCCTGAGCAATCCACTAAGAGGCTTGAGAGGCTTTGCTGCC 1312
DB |||||
QY 375 Val-----LeuPro-----ArgGlyLeuAsnGlyPhe----- 383
QY 1313 ACCTGGTGTGAGCCCTCAGTCACTTTCTCTGAATGCTTCTGAGCCT 1357
DB |||||
QY 394 ThrTrpAla-----SerSerThrAlaSerHisPro 393

RESULT 11
AA72638
ID AA72638 standard; protein; 395 AA.
XX
AC AA72638;
XX
DT 02-MAY-2001 (first entry)
XX
DE Mouse glycosyl sulfotransferase-4 (GST-4).
XX
KW Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy;
KW selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 8E1.
XX
OS Mus musculus.
XX
PN WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US019741.
XX
XX 20-JUL-1999; 99US-0144694P.
XX
XX 13-JUN-2000; 2000US-00593828.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX
XX MPI; 2001-138471/14.
XX
XX N-PSDB; AAD02696.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications.
XX
XX Claim 3; Fig 2; 128pp; English.
XX
XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
XX gene is found on chromosome 8E1. GST is a type 2 membrane protein useful
XX for inhibiting a binding event between a selectin and a selectin ligand,
XX which comprises contacting the selectin with a non-sulphated selectin
XX ligand, GST and a small molecular agent that inhibits the sulphation
XX activity of GST. GST is also useful in inhibiting a selectin mediated
XX binding event. GST is useful in gene therapy to treat disorders such as
XX acute or chronic inflammation, systemic lupus erythematosus (SLE),
XX rheumatoid arthritis, polyarthritis nodosa, polymyositis,
XX dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
XX myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
XX disease, adrenalitis, hypoparathyroidism, pernicious anaemia,
XX demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,
XX myocarditis, regional enteritis, adult respiratory distress syndrome,
XX infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial
XX asthma, hypersensitivity, rheumatic fever and tissue rejection during
XX transplantation
XX
XX SQ Sequence 395 AA;
XX
XX Alignment Scores:
XX
XX Pred. No.: 7,678-82 Length: 395
XX Score: 1020.00 Matches: 221
XX Percent Similarity: 66.58% Conservative: 44
XX Best Local Similarity: 55.53% Mismatches: 120
XX Query Match: 27.26% Indels: 13
XX DB: 4 Gaps: 8
XX
XX US-09-645-078-1 (1-2043) x AA72638 (1-395)
QY 128 ATGCTACTCCCTAAA-----AAAATGAGCTCTGCTGTTCTGTTCCAGATGCC 181
DB |||||
DB 1 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuLeuMetValGlnThrGly 20
QY 182 ATCTGGCTCTATTCTTCCACATGTACAGCCACACATCAGTCCCTCTCTATGAAGCA 241
DB |||||
DB 21 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 36
QY 242 CAGCCCGAGCGCATCAGTGTGTTCTGTTCTCTGCGCTGCTGCTCTCTTTTGTG 301
DB |||||
DB 37 LeuGlyGluArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheVal 56
QY 302 GGGCAGCTTTTGGCAGACCCAGATGTTTCTACTGATGAGCCCGCTGGCACGTG 361
DB |||||
DB 57 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal 76

QY 362 TGGATGACCTTCAAGCAGACCCGCTGGATGCTGCACATGGCTGTGGGATCTGATA 421
Db 77 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHsMetAlaValArgAspLeu 96
QY 422 CGGGCGCTTCTTTGTCACATGAGCGCTTTTGATGCTCATATGGAACCTGGTCCCGG 481
Db 97 ArgSerValPheLeuCysAspMetAspValPheAspAlaTyrLeu--ProTrpArgArg 115
QY 482 AGACAGTCCAGCTCTTTTCAGTGGGAGAACACCCGGGCGCTGTGTTCGACCTCCCTGT 541
Db 116 AsnIleSerAspLeuPheGlnTrpAlaValSerA-GAlaLeuCysSerProProValCys 135
QY 542 GACATCATCCCAAGATGAATCATCCCGGGCTCACTGCAGGCTCTGTGCAGTCAA 601
Db 136 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLysProLeuCysAlaThr 155
QY 602 CAGCCCTTTGAGTGTGAGAGAGCGCTCCCGCTCTACAGCCAGCTGGTGTCTAAGAG 661
Db 156 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 175
QY 662 GTGCGCTTCTCAACCTGAGCCCTCTACCGCTGCTGAAAGACCCCTCCCTCAACCTG 721
Db 176 ValArgPhePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 195
QY 722 CATATCGTCACCTGTCGGACCCCGCGCGTTCCTGCTCCGAGAACGCAAAAG 781
Db 196 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAla 215
QY 782 GGAGATCTCATGATTCACAGTCGATGTGAGGGGCGACGACGAGCAAACTCAAGAG 841
Db 216 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal--Glu 234
QY 842 GAGGACCAACCTTACTATGTATGTCAGGTCTCTGCCAAGCCAGCTGAGATCTACAAG 901
Db 235 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 254
QY 902 ---ACCATCCAGTCTTGGCCAGGCGCTGCGAGACGCTACTGCTGTGGCTATGAG 958
Db 255 AlaAlaLeuHisLysProProPheLeuGlnAspArgTyrArgLeuValArgTyrGlu 274
QY 959 GACCTGGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGATTTCTGGGATGGAA 1018
Db 275 AspLeuAlaArgAspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGly 294
QY 1019 TTCTTCCCATCTTACAGCTGGGTGTCATACATCACCCGAGGACGAGGATGGT--- 1075
Db 295 LeuThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCACGCTTCCACAAATGTCAGGATGCCCTTAATGTCTCCAGGCTTGGCGC 1132
Db 315 ArgArgGluAlaPheLysThrThrSerArgAspAlaLeuSerValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTTCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGGCGATGCCATGAAT 1192
Db 335 HisThrLeuProPheAlaLysIleArgArgValGlnGluLeuCysGlyGlyAlaLeuGln 354
QY 1193 TTCTGGGTACCGCCACGTCAGATCTGACACAGACAGAAACCTGTGTGGATCTT 1252
Db 355 LeuLeuGlyTyrArgSerValHisSerGluLeuGlnArgAspLeuSerLeuAspLeu 374
QY 1253 CTGTCTACCT---GGACTGTCTCCCTGAGCAATCCACTAAGAGGGTTGAGAAG 1301
Db 375 Leu-LeuProArgGlyMetAspSerPheLysTrpAlaSerSerThrGluLys 391

RESULT 12
AAU11275

ID AAU11275 standard; protein; 395 AA.

XX AC

XX AAU11275;

DT 12-MAR-2002 (first entry)

DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.

XX Mouse; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
KW antinflammatory; antipsoriatic; antidiabetic; dermatological;
KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
XX Mus musculus.
OS
XX WO200185177-A1.
PN
XX 15-NOV-2001.
PD
XX 10-MAY-2001; 2001MO-US015452.
PF
XX 11-MAY-2000; 2000US-00569320.
PR
XX (BURN-) BURNHAM INST.
PA
XX Fukuda M, Yeh J, Hiraoka N;
XX WPI; 2002-075226/10.
XX N-PSDB; AAS16948.
DR
XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
PT L-selectin sulfotransferase-2 that directs expression of L-selectin
PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
PT GlcNAc 6-sulfotransferase.
XX
XX Claim 28; Fig 10; 98pp; English.
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated beta1,3-N-
CC acetylglucosaminyltransferase (beta1,3GNT) or an active fragment, where
CC beta1,3GNT directs expression of a MECA-79 antigen. The invention also
CC provides a method of treating or preventing an L-selectin-mediated
CC condition by reducing the expression or activity of a beta1,3GNT that
CC directs expression of a MECA-79 antigen. This can be done by
CC administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3GNT,
CC and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LSST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3GNT. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents mouse I-GlcNAc6ST
XX
SQ Sequence 395 AA;
Alignment Scores:
Pred. No.: 7,67e-82 Length: 395
Score: 1020.00 Matches: 221
Percent Similarity: 66.58% Conservative: 44
Best Local Similarity: 55.53% Mismatches: 120
Query Match: 27.26% Indels: 13
DB: Gaps: 8
US-09-645-078-1 (1-2043) x AAU11275 (1-395)
QY 128 ATGCTACTGCTTAA-----AAATGAGCTCTGCTGTTTCTGTTTCCAGATGGCC 181
Db 1 MetArgLeuProArgPheSerThrValMetLeuSerLeuLeuMetValGlnThrGly 20
QY 182 ATCTTGGCTCTATTCTTCCACATGTACAGCCACAAACATCAGCTCCCTGTCTATGAGGCA 241
Db 21 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 36

DR WFI; 2001-138471/14.
 DR N-PSDB; AAD02697, AAD02698, AAD02699.
 XX
 XX
 PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 XX
 XX
 PS Claim 3; Fig 1; 128pp; English.
 XX
 XX
 CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
 CC alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
 CC membrane protein useful for inhibiting a binding event between a selectin
 CC and a selectin ligand, which comprises contacting the selectin with a non
 CC C-sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation. Note: The present sequence is also shown in
 CC sequence listing (page no: 56) but lacks four nucleotides at its 3' end
 CC
 XX Sequence 390 AA;
 XX
 Alignment Scores:
 Pred. No.: 7,32e-81 Length: 390
 Score: 1009.00 Matches: 206
 Percent Similarity: 66.41% Conservative: 57
 Best Local Similarity: 52.02% Mismatches: 105
 Query Match: 26.96% Indels: 28
 DB: 4 Gaps: 7
 XX
 US-09-645-078-1 (1-2043) x AAY72639 (1-390)
 Qy 122 AGCACATGCTACTGCTGCTAAAAAATGAAGTCTGTGCTTTCTGTGTTCCAGATGGCC 181
 Db 12 ThrValLeuLeuAlaGlnThrThrCysLeuLeuLeuPheIleIleSerArgPro---30
 Qy 182 ATCTGGCTCTATTCTTCACATGTACAGCCACACATCAGCTCCCTGCTGTATGAAGGCA 241
 Db 31 -----GlyProSerProAlaGly 37
 Qy 242 CAGCCCGAGCGCATGCATGCTGGTGTCTGCTCTCCGCGCTCTGGCTCTTTTGTG 301
 Db 38 GlyCluAspArgValHisValLeuValLeuSerSerTArgSerGlySerSerPheLeu 57
 Qy 302 GGGCAGCTTTTGGGAGCACCAGATGTTTCTACCTGTATGAGACCGCGCTGGCAGTG 361
 Db 58 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaIleHisVal 77
 Qy 362 TGGATGACCTTCAAGCAGAGCACCAGCGCTGGATGTGCACATGGCTGTGCGGATCTGATA 421
 Db 78 TrpThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuMet 97
 Qy 422 CGGCCGCTCTTTGTGCGACATGACGCGTCTTTGTATGCTACATGAGAACCTGTGTCGCCG 481
 Db 98 ArgSerIlePheLeuCysAspMetAspValPheAspAlaTyrMet---ProGlnSerArg 116
 Qy 482 AGACAGTCCAGCTCTTTTCAGTGGGAGAACAGCCGGCGCTGTGTCTTCGACCTGCCTGT 541
 Db 117 AsnLeuSerAlaPhePheAsnTAlaThrSerArgAlaLeuCysSerProProAlaCys 136
 Qy 542 GACATCATCCGACAGATGAATCATCCCCCGGGCTCACTCAGCGCTCTGTGAGTCAA 601
 Db 137 SerAlaPheProArgGlyThrIleSerLysGlnAspValCysLysThrLeuCysThrArg 156
 Qy 602 CAGCCCTTTTGGGTGGTGGAGAGGCGCTCCGCGCTCCCTCAGCCAGCTGGTCTCTCAAGGAG 661

Search completed: May 6, 2004, 11:01:02
Job time : 113.5 secs

Qy

242 CAGCCGAGCGCATGCACGTCTGGTTCGTCTCCTGCGCCTCTGGCCTCTTTTG 301
:::
:::

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 11:05:49 ; Search time 93.5 Seconds
(without alignments)
12129.816 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 3742
Sequence: 1 gaattccattgtgttggta.....tgggattctctaaaaa 2043

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Dlelop 6.0 , Dlext 7.0

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 2281346

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09645078/runat_06052004_104601_21872/app_query.fasta_1.2183
-DB=Published Applications_AA -QFM=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALICN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09645078 @CGN 1.1.13 @runat_06052004_104601_21872
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	2038	54.5	386	9	US-09-816-825-2	Sequence 2, Appli
2	2038	54.5	386	13	US-10-007-362-1	Sequence 1, Appli
3	1984	53.0	386	14	US-10-427-631-11	Sequence 11, Appli
4	1030.5	27.5	418	9	US-09-927-602-5	Sequence 5, Appli
5	1028.5	27.5	395	9	US-09-927-602-2	Sequence 2, Appli
6	1028.5	27.5	395	12	US-10-258-080-5	Sequence 5, Appli
7	1017.5	27.2	395	9	US-09-927-602-3	Sequence 3, Appli
8	1009	27.0	390	9	US-09-927-602-4	Sequence 4, Appli
9	821	21.9	171	9	US-09-927-602-8	Sequence 8, Appli
10	598.5	16.0	483	14	US-10-212-933-2	Sequence 4, Appli
11	587.5	15.7	484	14	US-10-212-933-4	Sequence 4, Appli
12	587.5	15.7	531	9	US-09-833-790-255	Sequence 255, Appli
13	549	14.7	411	12	US-10-211-462-97	Sequence 97, Appli
14	549	14.7	411	14	US-10-021-860-128	Sequence 128, App
15	519.5	13.9	169	9	US-09-927-602-6	Sequence 6, Appli
16	511.5	13.7	169	9	US-09-927-602-7	Sequence 7, Appli
17	504.5	13.5	481	12	US-10-087-192-123	Sequence 123, App
18	488	13.0	479	12	US-10-087-192-126	Sequence 126, App
19	336	9.0	169	9	US-09-927-602-9	Sequence 9, Appli
20	335.5	9.0	179	9	US-09-927-602-10	Sequence 10, Appli
21	293	7.8	174	9	US-09-927-602-11	Sequence 11, Appli
22	207.5	5.5	19695	15	US-10-084-846A-3	Sequence 3, Appli
23	201	5.4	19695	15	US-10-084-846A-3	Sequence 3, Appli
24	198.5	5.4	19608	15	US-10-084-846A-8	Sequence 8, Appli
25	193	5.2	19662	15	US-10-084-846A-6	Sequence 6, Appli
26	187.5	5.0	19725	15	US-10-084-846A-4	Sequence 4, Appli
27	182.5	4.9	19723	15	US-10-084-846A-5	Sequence 5, Appli
28	178	4.8	324	12	US-10-425-114-58150	Sequence 58150, A
29	173	4.6	19608	15	US-10-084-846A-8	Sequence 8, Appli
30	172.5	4.7	19652	15	US-10-084-846A-7	Sequence 7, Appli
31	170.5	4.6	4123	14	US-10-213-509-5	Sequence 5, Appli
32	170.5	4.6	4219	15	US-10-085-198-2	Sequence 2, Appli
33	170.5	4.6	19723	15	US-10-084-846A-5	Sequence 5, Appli
34	161	4.4	19725	15	US-10-084-846A-4	Sequence 4, Appli
35	160.5	4.3	336	12	US-10-425-114-56601	Sequence 56601, A
36	156	4.2	19662	15	US-10-084-846A-6	Sequence 6, Appli
37	155	4.1	2080	15	US-10-353-690-36	Sequence 36, Appli
38	147	4.0	595	15	US-10-342-331-48	Sequence 48, Appli
39	147	4.0	595	15	US-10-342-331-50	Sequence 50, Appli
40	147	4.0	797	14	US-10-156-761-10907	Sequence 10907, A
41	147	4.0	822	15	US-10-342-331-49	Sequence 49, Appli
42	146.5	3.9	3170	14	US-10-184-844-249	Sequence 249, App
43	146.5	3.9	3170	14	US-10-184-844-249	Sequence 249, App
44	145	3.9	613	15	US-10-260-937-16	Sequence 16, Appli
45	145	3.9	1464	12	US-09-918-715-261	Sequence 261, App

ALIGNMENTS

RESULT 1
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Alignment Scores: 1.49e-157 Length: 386
Pred. No.: 1.49e-157 Length: 386

Score: 2038.00 Matches: 386
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 54.46% Indels: 0
 DB: 9 Gaps: 0

US-09-645-078-1 (1-2043) x US-09-816-825-2 (1-386)

QY 128 ATGCTACTGCTAAATAAAGAGCTCTGCTCTTCTGTTCTCCAGATGGCCTGTTG 187
 Db 1 MetLeuLeuProLysLysMetLysLeuLeuPheLeuValSerGlnMetAlaLeu 20
 QY 188 GCTCTATTCTCCAGATGACAGCCACACATCAGTCTCCCTGCTATGAGGACAGCC 247
 Db 21 AlaLeuPhePheHisMetTyrSerHisAgnLysSerLeuSerMetLysAlaGlnPro 40
 QY 248 GAGCGATGACGTGCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307
 Db 41 GluArgMetHisValLeuValLeuSerSerTyrPargSerGlySerPheValGlyGln 60
 QY 308 CTTTTTGGGAGCAGCAGATGTTTCTACCTGATGAGCCGCCCTGGCAGCTGTGGATG 367
 Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTyrHisValTyrMet 80
 QY 368 ACCTTCAAGCAGACCGCTGATGCTGACATGCTGCGGATCTGATACGGCC 427
 Db 81 ThrPheLysGlnSerThrAlaTyrMetLeuHisMetAlaValArgAspLeuAla 100
 QY 428 GTCTTCTGTGCGATGAGCGCTTTTGTATGCTTACATGAACTGCTCCCGGAGACAG 487
 Db 101 ValPheLeuCysAspMetSerValPheAspAlaTyrMetGluProGlyProArgGln 120
 QY 488 TCCAGCCTCTTTCAGTGGGAGAACAGCGGGCCCTGTGTCTGTCACCTGTGTGACATC 547
 Db 121 SerSerLeuPheGlnTyrGluSerSerArgAlaLeuCysSerAlaProAlaCysAspIle 140
 QY 548 ATCCCAAGATGAATCATCCCGGGCTCACTGCAGGCTCCTGTGCTGCTCAACAGCC 607
 Db 141 IleProGlnAspGluIleProArgAlaHisCysArgLeuLeuCysSerGlnGlnPro 160
 QY 608 TTTGAGTGTGAGAGGCTGCTGCTCTACAGCCAGCTGCTGCTCAAGAGGTGGC 667
 Db 161 PheGluValValGluLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
 QY 668 TTCTTCAACCTGAGTCCCTCTACCGCTGCTGAAAGACCCCTCCCTCAACCTGCATATC 727
 Db 181 PhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsnLeuHisIle 200
 QY 728 GTGACCTGTGCGGAGACCCCGGGCTGCTTCCGAGAACGACAAAGGAGAT 787
 Db 201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAsp 220
 QY 788 CTGATGATTCAGCTGCTGATGGGAGCAGCATGAGCAAACTCAAGAGGAGGAC 847
 Db 221 LeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysGluAsp 240
 QY 848 CAACCTTACTATGTGATGAGTCACTGCTGCAAGCAGCTGAGATCTCAAGACATC 907
 Db 241 GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
 QY 908 CAGTCTTGGCCAGGCTGACGAGACGCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 967
 Db 261 GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla 280
 QY 968 CGAGCCCTGTGGCCAGACTTCCGGAATGATGAATTCCTGGATGGATTTCTTGGCC 1027
 Db 281 ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro 300
 QY 1028 CATCTTCAGACCTGGGTCATACATCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1087
 Db 301 HisLeuGlnThrTyrPheValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
 QY 1088 CACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTGGCGCTGCTTTCGCCCTAT 1147

Db 321 HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTyrArgTyrSerLeuProTyr 340
 QY 1148 GAAAGAGTCTTCTGACTTCAGAAAGCTGCTGGGATGCCATGAATTTCTGGGCTACCGC 1207
 Db 341 GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrArg 360
 QY 1208 CACCTCAGATCTGAACAAGAACAGAGAACTCTTCTGATCTTCTGCTTACTGGAAT 1267
 Db 361 HisValArgSerGluGlnGluArgAsnLeuLeuAspLeuLeuSerThrTyrThr 380
 QY 1268 GTCCCTGAGCAATCCAC 1285
 Db 381 ValProGluGlnIleHis 386

RESULT 2
 US-10-007-262-1
 ; Sequence 1, Application US/10007262
 ; Publication No. US20020164748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/10/007,262
 ; PRIORITY FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 ; US-10-007-262-1

Alignment Scores:
 Pred. No.: 1,49e-157 Length: 386
 Score: 2038.00 Matches: 386
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 54.46% Indels: 0
 DB: 13 Gaps: 0

US-09-645-078-1 (1-2043) x US-10-007-262-1 (1-386)

QY 128 ATGCTACTGCTAAATAAAGAGCTCTGCTGTTCTGTTCTCCAGATGGCCTGTTG 187
 Db 1 MetLeuLeuProLysLysMetLysLeuLeuPheLeuValSerGlnMetAlaLeu 20
 QY 188 GCTCTATTCTCCACATGTACAGCCACACATCAGTCCCTGCTATGAAAGGACAGCC 247
 Db 21 AlaLeuPhePheHisMetTyrSerHisAgnLysSerLeuSerMetLysAlaGlnPro 40
 QY 248 GAGCGATGACGTGCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307
 Db 41 GluArgMetHisValLeuValLeuSerSerTyrPargSerGlySerPheValGlyGln 60
 QY 308 CTTTTTGGGAGCAGCAGATGTTTCTACCTGATGAGCCGCCCTGGCAGCTGTGGATG 367
 Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTyrHisValTyrMet 80
 QY 368 ACCTTCAAGCAGACCGCTGATGCTGACATGCTGCGGATCTGATACGGCC 427
 Db 81 ThrPheLysGlnSerThrAlaTyrMetLeuHisMetAlaValArgAspLeuAla 100
 QY 428 GTCTTCTGTGCGATGAGCGCTTTTGTATGCTTACATGAACTGCTCCCGGAGACAG 487
 Db 101 ValPheLeuCysAspMetSerValPheAspAlaTyrMetGluProGlyProArgGln 120
 QY 488 TCCAGCCTCTTTCAGTGGGAGAACAGCGGGCCCTGTGTCTGTCACCTGTGTGACATC 547


```
Db 121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuCysSerAlaProAlaCysAspIle 140
Qy 548 ATCCCAAGATGAATCATCCCGGGCTCACTGCAGGCTCTGTGCTCAACAGCCC 607
Db 141 IleProGlnAspGluIleProArgAlaHisCysArgLeuLeuCysSerGlnGlnPro 160
Qy 608 TTGAGGTGTGTGAGAGGCTGCGCTCTACAGCCACGCTGCTCAAGAGGTCGC 667
Db 161 PheGluValValGluLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
Qy 668 TTCTTCAACCTGAGTCCCTTACCGCTGCTGAAGACCCCTCCCTCAACCTGCATC 727
Db 181 PhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsnLeuHisIle 200
Qy 728 GTGCACCTGTGTCCCGGACCCCGGGCTGTTCCTGCTCCGAGAACCCACAAAGGAGAT 787
Db 201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAsp 220
Qy 788 CTCATGATTGACGTGCGATTGTGATGGGCGAGCATGAGCAAAACCTCAAGAGGAGAC 847
Db 221 LeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAsp 240
Qy 848 CAACCTCTACTATGTGATCAGGTCTATCTGCCAAAGCCAGCTGGAGATCTCAAGACCATC 907
Db 241 GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
Qy 908 CAGTCTTGGCCCAAGCCCTGAGGAAGCTACCTGCTGTGTGCGCTATGAGGACCTGCT 967
Db 261 GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValAlaGlyrGluAspLeuAla 280
Qy 968 CGAGCCCTGTGGCCCGACACTTCCCGAATGTATGAATTCGTGGGATTGGAAATCTTGCCC 1027
Db 281 ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro 300
Qy 1028 CATCTTCAGACCTGGGTGATTAACATCACCCGAGGCGATGGGTGACACGCTTTC 1087
Db 301 HisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
Qy 1088 CACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGTGTCTTTCCTTAT 1147
Db 321 HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTyr 340
Qy 1148 GAAAGGTTTTCGACTTCAGAAAGCCTGTGGCGATGCCATGAATTTGCTGGGTACCGC 1207
Db 341 GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuGlyTyrArg 360
Qy 1208 CACGTCAGATCTGAACAGAACAGAGAACCTGTGCTGCTGCTTCTCTACCTGGACT 1267
Db 361 HisValArgSerGluGlnGluGlnArgAsnLeuLeuLeuAspLeuLeuSerThrTrpThr 380
Qy 1268 GTCCCTGAGCAATCCAC 1285
Db 381 ValProGluGlnIleHis 386
```

```
RESULT 3
US-10-427-631-11
; Sequence 11, Application US/10427631
; Publication No. US20030175923A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: CORLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT APPLICATION NUMBER: US/10/427,631
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
```

```
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: m3sc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
; US-10-427-631-11

Alignment Scores:
Pred. No.: 386e-153 Length: 386
Score: 1984.00 Matches: 383
Percent Similarity: 99.48% Conservative: 2
Best Local Similarity: 98.97% Mismatches: 1
Query Match: 53.02% Indels: 2
DB: 14 Gaps: 0

US-09-645-078-1 (1-2043) x US-10-427-631-11 (1-386)
Qy 128 ATGCTACTCCCTAAATAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGCCCATTTG 187
Db 1 MetLeuLeuProLysLysMetLysLeuLeuPheLeuValSerGlnMetAlaIleLeu 20
Qy 188 GCTCTATTCTCCACATGTACAGCCACACATCAGCTCCCTGCTCTATGAGGCACAGCCC 247
Db 21 AlaLeuPhePheHisMetTyrSerHisAsnIleSerSerLeuSerMetLysAlaGlnPro 40
Qy 248 GAGCGATGACGCTGCTGCTGCTGCTCTCTCTGCGCTCTGCTCTCTCTTTTGTGGGCGAG 307
Db 41 GluArgMetHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGlyGln 60
Qy 308 CTTTTTGGGACAGCCACCATGTTTCTACGTATGAGAGCCCGCTGCGGATCTGATACGGGC 367
Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrpMet 80
Qy 368 ACCTTCAAGCAGACAGCCCGCTGGATGTCGATGTCGCTGTCGGGATCTGATACGGGC 427
Db 81 ThrPheLysGlnSerThrAlaTrpMetLeuHisMetAlaValArgAspLeuIleArgAla 100
Qy 428 GTCTTCTTGTGCGACATGAGCGCTTTTGTAGTCCCTACATGGAACCTGCTCCCGGAGACAG 487
Db 101 ValPheLeuCysAspMetSerValPheAspAlaTyrMetGluProGlyProArgArgGln 120
Qy 488 TCCAGCTCTTTTCAGTGGGAGACACCGCGGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 547
Db 121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuCysSerAlaProAlaCysAspIle 140
Qy 548 ATCCCAAGATGAATCATCCCGGGCTCACTGCAGGCTCCTGTCGATCAACAGCCC 607
Db 141 IleProGlnAspGluSerSerProGlyLeuThrAlaGlySerCysAlaValAsnSerPr 160
Qy 608 TTGAGGTG-GTGAGAGAGCCCTGCGCTCTCAACAGCCAGCTGCTGCTCAAGGAGGTGCG 666
Db 160 oLeuLysLeuLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
Qy 667 CTTCTTCAACCTGAGTCCCTCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726
Db 180 gPhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsnLeuHisI 200
Qy 727 CGTGCACCTGCTGGGAGACCCCGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 786
Db 200 eValHisLeuValArgAspProArgAlaValPheArgSerArgGlnArgThrLysGlyAs 220
Qy 787 TCTCATGATTGACAGTCCGCTTGTGATGGGCGAGCATGAGCAAAATCTCAAGAGGAGGA 846
```

Db 220 pLeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAs 240
QY 847 CCAACCTACTATGATCAGGTCTATCTGCCAAAGCCAGCTGGAGATCTCAAGACCAT 906
Db 240 pGlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrI 260
QY 907 CCAGTCTTCCCAAGCCCTCAGGAAGCTACCTGCTTGTGCGCTATGAGGACCTGGC 966
Db 260 eGlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAl 280
QY 967 TCAGAGCCCTGTGGCCCAAGCTTCCCGAATGATGAATTCGTGGGATTGGAATTCCTGCC 1026
Db 280 aargAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPr 300
QY 1027 CCATCTTCAGACTGGTGCATAAATCACCAGGAGGAGGATGGGTGACCAAGCTTT 1086
Db 300 ohisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaph 320
QY 1087 CCACAAATGCCAGGATGCCCTTAATGCTCCAGAGCTGGCGCTGCTTTCGCCCTA 1146
Db 320 ehisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTpaGTrpSerLeuProTy 340
QY 1147 TGAAGGTTTCTGACTCTCAGAAAGCTGTGGCATGCCATGAATTCCTGGGCTACG 1206
Db 340 rGluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrAr 360
QY 1207 CCAGTTCAGATCTGAACAAGAACAGAGAAACCTGTGCTGGATCTCTCTACCTGGAC 1266
Db 360 gHisValArgSerGluGlnGluGlnArgAsnLeuLeuAspLeuLeuSerThrTrpTh 380
QY 1267 TGTCCCTCAGCAATCCAC 1285
Db 380 rValProGluGlnIleHis 386

RESULT 4

US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

Alignment Scores:
Pred. No.: 2,99e-75 Length: 418
Score: 1030.50 Matches: 221
Percent Similarity: 66.75% Conservative: 44
Best Local Similarity: 55.67% Mismatches: 120
Query Match: 27.54% Indels: 12
DB: 9 Gaps: 7

US-09-645-078-1 (1-2043) x US-09-927-602-5 (1-418)

QY 128 ATGCTACTGCTAAA-----AAATGAAGCTCTGCTGTTCTGTTTCCCGATGGCC 181
Db 25 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuLeuMetValGlnThrGly 44
QY 182 ATCTTGGCTCTATTCTTCCACATCATAGCCCAACATCAGCTCCCTGTCTATGAAGGCA 241
Db 45 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 60

QY 242 CACCCGAGCGCATGACGCTGCTGTTCTCTCTCTCTGCGCTGCTGCTCTCTCTTTTG 301
Db 61 LeuGlyGluArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheVal 80
QY 302 GGGCAGCTTTTGGGAGCAGCAGCCAGATGTTTCTACCTGATGAGCCCGCTGGCAGCG 361
Db 81 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal 100
QY 362 TGGATGACCTTCAAGCAGAGCAGCCGCTGGATGCTGCACATGCTGCGGATCTGATA 421
Db 101 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeuIle 120
QY 422 CCGGCGCTCTCTGTCGACATGAGCTCTTGTGATGCTTACATGAACTGCTGCTCCCG 481
Db 121 ArgSerValPheLeuLeuCysAspMetAspValPheAspAlaTyrLeu---ProTrpArg 139
QY 482 AGACAGTCCAGCTCTTTCAGTGGGAGAACAGCCGCGCTCTGTTCTGCACCTGCCCT 541
Db 140 AsnIleSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProValCys 159
QY 542 GACATCATCCCAAGATGAATCATCCCGCGCTCACTGAGGCTCTCTGTCGATCAA 601
Db 160 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLysProLeuCysAlaThr 179
QY 602 CAGCCCTTTGAGTGTGGAGAGGCTCGCGCTCTCTACAGCCACGCTGCTGCTCAAGGAG 661
Db 180 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 199
QY 662 GTGCGCTCTTCAACTGACCTGCTCTACCCGCTCTGAAAGACCCCTCCCTCAACCTG 721
Db 200 ValArgPhePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 219
QY 722 CATATCGTCACCTGCTCGGAGCCCGCGCGCTGTTCCGTTCCGAGAACGCAAG 781
Db 220 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAla 239
QY 782 GGAGATCTCATGATTGACAGCTGATGGGAGCAGCATGAGCAAAACATCAAGAAG 841
Db 240 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---Glu 258
QY 842 GAGGACCAACCTACTATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 901
Db 259 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 278
QY 902 ACCATCAGCTCTTCCCAAGGCTCGAGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
Db 279 AlaLeuHisLysProProProPheLeuGlnAspArgTyrArgLeuValArgTyrGluAsp 298
QY 962 CTGCTCGAGCCCTGTGGCCAGACTTCCCGAATGATGAATTCGTTGGGATTGGAATTC 1021
Db 299 LeuAlaArgAspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGlyLeu 318
QY 1022 TTGCCCCATCTTTCAGACTGGTGTCATACATCACCGAGGCAAGGCGCTGGT----- 1075
Db 319 ThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAlaArg 338
QY 1076 GACCACGCTTTCCACACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTTGGCGTGG 1135
Db 339 ArgGluAlaPheLysThrThrSerArgAspAlaLeuSerValSerGlnAlaTrpArgHis 358
QY 1136 TCTTTGCCCTATGAAAGTTCGACTTCAGAACCCCTGCGGATGCGGATGCGGATTTG 1195
Db 359 ThrLeuProPheAlaLysIleArgArgValGlnGluLeuCysGlyGlyAlaLeuGlnLeu 378
QY 1196 CTGGGCTACCCCGCAGCTCAGATCTGAACAAGAACAGAGAAACCTGTTGCTGGATCTTCTG 1255
Db 379 LeuGlyTyrArgSerValHisSerGluLeuGlnArgAspLeuSerLeuAspLeuLeu 398
QY 1256 TCTACCT---CGACTGCTCCCTGAGCAAAATCCACTAAGAGGTTCCAGAG 1301
Db 399 -LeuProArgGlyMetAspSerPheLysTrpAlaSerSerThrGluLys 414

[illegible]

```

; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/202,234
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/200,185
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/198,403
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/USC1/11869
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/197,590
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID NO. US20040029125A1 7472777CD1
US-10-258-080-5

Alignment Scores:
Pred. No.: 4,288-75 Length: 395
Score: 1028.50 Matches: 224
Percent Similarity: 65.06% Conservative: 46
Best Local Similarity: 53.98% Mismatches: 108
Query Match: 27.49% Indels: 37
Dbs: 12 Gaps: 10

US-09-645-078-1 (1-2043) x US-10-258-080-5 (1-395)
QY 125 ACAATGCTACTGCTCCCTAAATAAGCTCCTGCTGTTCTGCTGTTCCAGATGCCATC 184
Db 12 ThrAlaLeuLeuAlaGlnThrPheLeuLeuLeuPheLeuValSerArgPro----- 29
QY 185 TTGGCTCTATTCTTCCACATGATACACCAACATCAGTCCCTGCTCTATGAAGCCACAG 244
Db 30 -----GlyProSerSerProAlaGlyGly 37
QY 245 CCCGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 304
Db 38 GluAlaArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGly 57
QY 305 CAGCTTTTGGCAGACCCAGATGTTTCTACCTGATGGAGCCGCTGCGCATCGG 364
Db 58 GlnLeuPheAsnGlnHisProAspValPheValLeuMetGluProAlaTrpHisValTrp 77
QY 365 ATGACCTTCAAGCAGACACCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 424
Db 78 ThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuValArg 97
QY 425 GCGGTCTTCTGCGACATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484
Db 98 SerValPheLeuCysAspMetAspValPheAspAlaTrpLeu---ProTrpArgArgAsn 116
QY 485 CAGTCCAGCTCTTTTCAGTGGAGAACACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 544
Db 117 LeuSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProAlaCysSer 136
QY 545 ATCATCCACAGATGAATATCATCCCGGCTCAGTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 604
Db 137 AlaPheProArgGlyAlaIleSerSerGluAlaValCysLysPheLeuCysAlaArgGln 156
QY 605 CCTTTGAGTGTGGAGAGGCTCCGCTCCTACAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 664
Db 157 SerPheThrLeuAlaArgGluAlaCysArgSerTyrSerHisValValLeuLysGluVal 176
QY 665 CGTCTTCTAACCTGAGTCCCTCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 724
Db 177 ArgPhePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeuArg 196
QY 725 ATCGTGACCTGCTGCGGAGACCCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 784

```

```

Db 197 IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys 216
QY 785 GATCTCATGTTGACATCGCATTTGTATGGGCGCAGCATGAGCAAAACCAAGCAAGAG 844
Db 217 AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla 235
QY 845 GACCAACCTTACTATGTCATGTCATCTGCCAAAGCCAGCTGGAGATCTCAAG--- 901
Db 236 AspProGlyLeuArgValValArgGluValCysArgSerHisValArgIleAlaGluAla 255
QY 902 ---ACATCCAGCTCTGCCCAAGCCCTGCGAGGACCTACCTGCTGCTGCTGCTGCTGCTG 958
Db 256 AlaThrLeuLysPro---ProPheLeuArgGlyArgTyrArgLeuValArgPheGlu 274
QY 959 GACCTGGCTCGAGCCCTGTCGCCAGACTTCCCGAATGTATGAATGCTGGGATTTGAA 1018
Db 275 AspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThrGlyLeuSer 294
QY 1019 TTCCTGGCCCATCTTCAGACCTGGTGTGATAACATCACCCGAGCAAGGCGATGGT--- 1075
Db 295 LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCACGCTTTCACACAAATCCAGGGATGCCCTTAATGTCTCCAGGCTTGGCGC 1132
Db 315 ArgArgGluAlaPheLysThrSerSerArgAsnAlaLeuAsnValSerGlnAlaTrpArg 334
QY 1133 TGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1192
Db 335 HisAlaLeuProPheAlaLysIleArgValGlnGluLeuCysAlaGlyAlaLeuGln 354
QY 1193 TTGCTGGGCTACCGCCACGTCAGATCTGAACAGAACAGAGAAACCTGTGTGCTGATCTT 1252
Db 355 LeuLeuGlyTyrArgProValTyrSerGluAspGluGlnArgAsnLeuAlaLeuAspLeu 374
QY 1253 CTGCTTACCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1312
Db 375 Val-----LeuPro-----ArgGlyLeuAsnGlyPhe----- 383
QY 1313 ACCTGCTGTCAGCTCAGTCACTTCTCTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1357
Db 384 ThrTrpAla-----SerSerThrAlaSerHisPro 393

RESULT 7
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; DYSTROPHY
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; NAME/KEY: VARIANT
; LOCATION: (1)...(395)
; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Alignment Scores:
Pred. No.: 3,39e-74 Length: 395
Score: 1017.50 Matches: 213

```

Percent Similarity:	67.72%	Conservative:	45
Best Local Similarity:	55.91%	Mismatches:	102
Query Match:	27.19%	Indels:	21
DB:	9	Gaps:	6

US-09-645-078-1 (1-2043) X US-09-927-602-3 (1-395)

QY	125	ACAATGCTACTGCTCTAAAAAATGAAGCTCTGCTGCTTTCTGTTTCCAGATGGCCATC	184
Db	12	Thr***LeuLeuAlaGlnThr***LeuLeuPheLeuValSerArgPro-----	29
QY	185	TTGGCTCTATTCTTCCACATGTACAGCCACAATCAGCTCCCTGTCTATGAAGGCACAG	244
Db	30	-----GlyProSerSerProAlaGlyGly	37
QY	245	CCGAGCGGANGCACTGCTGTTCTGTCTTCTGCGCTCTGCTCTCTCTTTTGTGGG	304
Db	38	Glu***ArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGly	57
QY	305	CAGCTTTTGGCAGCACCCAGATGTTTCTACTCATGTGAGCCCGCTGGCAGCTGTGG	364
Db	58	GlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrp	77
QY	365	ATGACCTTCAAGCAGACGACCGCTGGATGCTGCACATGCTGTGGGGATCTGTATACGG	424
Db	78	ThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeu***Arg	97
QY	425	GGCGTCTTCTTGGCAGCATGACGCTCTTTGATGCTCATAGGAACCTGGTCCCCGGAGA	484
Db	98	SerValPheLeuCysAspMetAspValPheAspAlaTyrLeu---ProTrpArgArgAsn	116
QY	485	CAGTCCAGGCTCTTTCAGTGGGAGAACAGCGCGGCGCTGTGTCTGCACCTGCTGTGAC	544
Db	117	LeuSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProAlaCysSer	136
QY	545	ATCATCCCAAGATGAATCATCCCGGGCTCATCTGAGGCTCTCTGTGCACTCAACAG	604
Db	137	AlaPheProArgGly***IleSerSerGlu***ValCysLysProLeuCysAlaArgGln	156
QY	605	CCCTTTGAGTGTGTGAGAGGCTCCGCTCCCTACAGCCACGCTGTGTCTCAGGAGCTG	664
Db	157	ProPhe***LeuAlaArgGluAlaCysGserTyrSerHisValValLeuLysGluVal	176
QY	665	CGCTTCTTCAACTGCAGTCCCTCTACCGCTGCTGAAAGACCCCTCCCTCAACCTGCAT	724
Db	177	ArgPhePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeuArg	196
QY	725	ATCGTCACCTGTGTCGGAGACCCCGCGGCGTGTTCGTTCCGACAGCACGCAAAAGGA	784
Db	197	IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys	216
QY	785	GATCTCATGATTACAGTTCGATTGTGATGGGCGCAGCATGACGAGCAAACTCAAGAAGGAG	844
Db	217	AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla	235
QY	845	GACCAACCTACTATGTGATGAGGTCTATCTGCCAAAGCCAGCTGGAGATCTACAAG---	901
Db	236	AspPro***LeuArgValValArgGluValCysArgSerHisValArgIleAlaGluAla	255
QY	902	---ACCATCCAGTCCCTTCCCAAGGCCCTCCAGAACGCTACCTGTGTGTCGCTATGAG	958
Db	256	AlaThrLeuLysPro---ProProPheLeuArgGlyArgTyrArgLeuValArgPheGlu	274
QY	959	GACCTGGCTGAGCCCTGTGGCCAGACTTCCCGAATGTATGAATTCGGGGATTGGAA	1018
Db	275	AspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThrGlyLeu***	294
QY	1019	TTCTTCCCATCTTTCAGACCTGGGTGCATAACATCACCGAGGACGGCATGGGT---	1075
Db	295	LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla	314
QY	1076	---GACCACGCTTTCCACACAAATGCCAGGGATGCCCTTAATGCTCTCCAGGCTTGGCGC	1132

```

315 ArgArgGluAlaPheLysThrSerSerArgAenAlaLeuAenValSerGlnAlaTrpArg 334
1133 TGGCTCTTGGCCCTATGAAAGGTTCTTCGACCTTCAGAAAGCCCTGGCGCATGCCATGAAT 1192
      ::::::::::::::::::::: 1133 :::::::::::::::::::::
335 HisAlaLeuProPheAlaLysIleArgArgValGlnGluLeuCysAlaGlyAlaLeuGln 354
1193 TTGCTGGGCTACCGCCAGCTCAGATCTGAACACAGAACACAGAGAAACCTGTTGCTGGATCTT 1252
      ::::::::::::::::::::: 1193 :::::::::::::::::::::
355 LeuLeuGlyTyrArgProValTyrSerGluAepGluGlnArgAspLeu***LeuAspLeu 374
1253 CTG 1255
      ::::
375 Val 375

RESULT 8
US-09-927-602-4
; Sequence 4, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Akana, Tomoya O.
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-4

Alignment Scores:
Pred. No.: 1,67e-73 Length: 390
Score: 1009.00 Matches: 296
Percent Similarity: 66.41% Conservative: 57
Best Local Similarity: 52.02% Mismatches: 105
Query Match: 26.96% Indels: 28
DB: 9 Gaps: 7

US-09-645-078-1 (1-2043) x US-09-927-602-4 (1-390)
QY 122 AGCAAAATGCTACTGCTAAATAAATGAAGCCTCTGCTGTTTCCAGATGGCC 181
      ::::::::::::::::::::: 122 :::::::::::::::::::::
DB 12 ThrValLeuLeuAlaGlnThrThrCysLeuLeuPheIleSerArgPro--- 30
      ::::::::::::::::::::: 182 APTCTGGCTCTATTCTTCCACATGACAGCCACAACATCAGCTCCCTGCTGTATGAAGCCA 241
      ::::::::::::::::::::: 31 -----GlyProSerSerProAlaGly 37
      ::::::::::::::::::::: 242 CAGCCGAGCGCATGCAGCTGCTGTTCTGCTCTCTGCGCGCTCTGCGCTCTCTTTTGTG 301
      ::::::::::::::::::::: 38 GlyGluAepArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheLeu 57
      ::::::::::::::::::::: 302 GGGCGAGCTTTTGGCGAGCACCAGATGTTTCTACCTCATGAGAGCCCGCTGGCACGTG 361
      ::::::::::::::::::::: 58 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal 77
      ::::::::::::::::::::: 362 TGGATGACCTTCAGCAGAGCACCCTGGATGCTGCACATGGCTGTGGGATCTGATA 421
      ::::::::::::::::::::: 78 TrpThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuMet 97
      ::::::::::::::::::::: 422 CGGGCCGCTCTCTTGTGGCAGCATGAGCGCTTTGTAGCCCTACATGGAACTGTCCTCCCGG 481
      ::::::::::::::::::::: 98 ArgSerIlePheLeuCysAspMetAspValPheAspAlaTyrMet---ProGlnSerArg 116
      ::::::::::::::::::::: 482 AGACAGTCCAGCCTCTTTCAGTGGGAGAACAGCCGGCCCTGTGTTCTGCACTGCTGT 541
      ::::::::::::::::::::: 117 AsnLeuSerAlaPhePheAsnTrpAlaThrSerArgAlaLeuCysSerProProAlaCys 136
      :::::::::::::::::::::
DB 117 AsnLeuSerAlaPhePheAsnTrpAlaThrSerArgAlaLeuCysSerProProAlaCys 136

```

QY 542 GACATCATCCCAAGATGAATCATCCCGGGCTCACTGAGGCTCCTGTGAGTCAA 601
Db 137 SerAlaPheProArgGlyThrIleSerIysGlnAspValCysIleThrLeuCysThrArg 156
QY 602 CAGCCCTTTGAGTGGTGGAGAGCCCTCCCTCTACAGCCAGCTGCTCTCAAGGAG 661
Db 157 GlnProPheSerLeuAlaArgGluAlaCysArgSerTyrSerHisValValLeuIysGlu 176
QY 662 GTGCGCTTCTTCAACCTGAGTCCCTTACCCGCTGCTGAAAGCCCTCCCTCAACCTG 721
Db 177 ValArgPhePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 196
QY 722 CATATGTGACCTGTCGGGACCCCGGCGCTGTCCTCCGTCCTCCGAGAACGACAAAG 781
Db 197 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluAlaAlaGly 216
QY 782 GGAGATCTCATGATTCAGATCGCATTTGATGGGCGCAGCATGACCAAACTCAAGAG 841
Db 217 ProIleLeuAlaArgAspAsnGlyIleValLeuGlyThrAsn---GlyIleStrpValGlu 235
QY 842 GAGGACCAACCTTACTATGTGATGATGAGTCACTCCCAAGCCAGCTGGAGATCTACAAG 901
Db 236 AlaAspProHisLeuArgLeuIleArgGluValCysArgSerHisValArgIleAlaGlu 255
QY 902 -----ACCATCCAGTCTTCCCAAGCCCTGCGAGGACGCTACTCTGTGGCTAT 955
Db 256 AlaAlaThrLeuIlePro---ProProPheLeuArgGlyArgTyrArgLeuValArgPhe 274
QY 956 GAGGACCTGGCTCGAGCCCTGTCGCGAGACTTCCGAATGTATGAATTCGTGGGATTG 1015
Db 275 GluAspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThrGlyLeu 294
QY 1016 GAATCTTGGCCCATCTTCAGACCTGGGTGTCATACATCACCCGAGGCAAGGATGGGT 1075
Db 295 ThrLeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyIleGly 314
QY 1076 GAC-----CAGCCTTTCACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTGG 1129
Db 315 LysProIleGluAlaPheHisThrSerSerArgAsnAlaArgAsnValSerGlnAlaTrp 334
QY 1130 CCGTGGTCTTCCCTATGAAAGTTCTCGACTTCAGAAAGCTGTCGGCGATGCCATG 1189
Db 335 ArgHisAlaLeuProPheThrLysIleLeuArgValGlnGluValCysAlaGlyAlaLeu 354
QY 1190 AATTGTCTGGGTACCGCCACCTCAGATCTGAACAAGAACAGAAACCTGTGTGGAT 1249
Db 355 GlnLeuLeuGlyTyrArgProValTyrSerAlaAspGlnGlnArgAspLeuThrLeuAsp 374
QY 1250 CTTCTG-----TCTACTGGACTGTCCCTGAG 1276
Db 375 LeuValLeuProArgGlyProAspHisPheSerTrpAlaSerProAsp 390

RESULT 9

US-09-927-602-8
; Sequence 8: Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dys trophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo Sapien

US-09-927-602-8

Alignment Scores: 2,91e-58 Length: 171
Pred. No.: 821.00 Matches: 171
Score: 60.00% Conservative: 0
Best Local Similarity: 60.00% Mismatches: 0
Query Match: 21.94% Indels: 114
DB: 9 Gaps: 2

US-09-645-078-1 (1-2043) x US-09-927-602-8 (1-171)

QY 221 AGTCTCCTGTCATGAAAGCAGCAGCCGAGCATGACGTCGTCTGTTCTGTTCTCTGG 280
Db 1 SerSerLeuSerMetIysAlaGlnProGluArgMetHisValLeuValLeuSerTyr 20
QY 281 CGCTCTGGGCTCTTCTTTCTGTTGGGAGCTTTTGGCAGCACCAGATGTTTCTACCTG 340
Db 21 ArgSerGlySerSerPheValGlnLeuPheGlyGlnHisProAspValPheTyrLeu 40
QY 341 ATGAGACCCGCTTGGCAGCTGTGGATGACCTTCAAGCAGAGCACCCTGGATGCTGCAC 400
Db 41 MetGluProAlaTrpHisValTrpMetThrPheLys----- 52
QY 401 ATGCTGTGCGGATCTGATACGGGCCGCTCTTCTGTGCGACATGAGCGTCTTTGATGCC 460
Db 52 ----- 52
QY 461 TACATGGAACCTGGTCCCCGAGACAGTCCAGCCTCTTTTCAGTGGGAGAACAGCCGGGCC 520
Db 52 ----- 52
QY 521 CTGTGTTCTGACCTGCTGTCATCATCCCAAGATGAAATCATCCCCGGGCTCAC 580
Db 52 ----- 52
QY 581 TGCAGGCTCTCTGTGCAGTCAACAGCCCTTTGAGTGGTGGAGAGGCCCTGCGCTCTAC 640
Db 53 -----LysAlaCysArgSerTyr 58
QY 641 AGCCAGCTGTGCTCAAGAGGTGGCTCTTCAACCTGAGTCCCTCTACCCGCTGCTG 700
Db 59 SerHisValValLeuIysGluValArgPhePheAsnLeuGlnSerLeuTyrProLeuLeu 78
QY 701 AAAGACCCCTCCCTCAACCTGCATATCGTGCACCTGTCGGGAGCCCGGCGCTGTTC 760
Db 79 LysAspProSerLeuAsnLeuHisIleValHisLeuValArgAspProArgAlaValPhe 98
QY 761 CGTTCCCGAGAACGCAAAAGGAGATCTCATGATTCAGATTCGATTCGATGGGCGAG 820
Db 99 ArgSerArgGluArgThrLysGlyAspLeuMetIleAspSer----- 112
QY 821 CATGAGCAAAACTCAAGAAGGAGGAGCAACCTTACTATGTGATGAGGTCTATCTGCCAA 880
Db 112 ----- 112
QY 881 AGCCAGCTGGAGATCTTACAGACCATTCAGTCTCTTGGCCCAAGCCCTGAGAAACGCTAC 940
Db 113 -----LysThrIleGlnSerLeuProLysAlaLeuGlnGluArgTyr 126
QY 941 CTGCTTGTGGCTATGAGGACCTGGCTGAGCCCTGTCGAGCCCTGTCGGCCAGACTTCCCGAATGTAT 1000
Db 127 LeuLeuValArgTyrGluAspLeuAlaArgAlaProValAlaGlnThrSerArgMetTyr 146
QY 1001 GAATCTGTGGATTTGGAATTTCTTGGCCCATCTTCAGACCTGGGTGCATAACATCACCGCA 1060
Db 147 GluPheValGlyLeuGluPheLeuProHisLeuGlnThrTrpValHisAsnIleThrArg 166
QY 1061 GCGAAGGCGATGGGT 1075
Db 167 GlyLysGlyMetGly 171

RESULT 10


```

Db      263 -----LysProTyrAsnLeuAspValThrGlnLeuThrThrValCysGluAsp 278
QY      884 CAGCTGGAGATCTACAAGACCATCCAGTCTCTGCGCCAGGCCCTGCAGGAACGCTACCTG 943
Db      279 PheSerAsnSerValSerThrGlyLeuMetArgProProtrpLeuLysGlyLysTyrMet 298
QY      944 CTTGTGGCTATGAGGACCTGCTGAGCCCTGAGCCCTGCGGACGACTTCCCGAATGTATGAA 1003
Db      299 LeuValArgTyrGluAspLeuAlaArgAsnProMetLysLysThrGluGluIleTyrGly 318
QY      1004 TTCTGGGATGGAAATCTTCTCCCATCTTCAGACCTGGGTGCATACATCACCCGAGGC 1063
Db      319 PheLeuGlyIleProLeuAspSerHisValAlaArgTyrPheGlnAsnAsnThrArgGly 338
QY      1064 ---AAGGCATGGTGACCGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCC 1120
Db      339 AspProThrLeuGlyLysHisLysTyrGlyThr---ValArgAsnSerAlaAlaThrAla 357
QY      1121 CAGCCTGGCCTGGTCTTCTCCCTATGAAAGCTTTCGACTTCAGAAAGCCTGTGGC 1180
Db      358 GluLysTrpArgPheArgLeuSerTyrAspIleValAlaPheAlaGlnAsnAlaCysGln 377
QY      1181 GATGCCATGAATTTGCTGGCTACGCCACGTCAGATCTGTGACAAAGACAGAAACCTG 1240
Db      378 GlnValLeuAlaGlnLeuGlyTyrLysIleAlaAlaSerGluGluLeuLysAsnPro 397
QY      1241 TTGCTGGATCTCTG 1255
Db      398 SerValSerLeuVal 402

RESULT 14
US-10-021-660-128
; Sequence 128, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1 Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-00071005
; CURRENT APPLICATION NUMBER: US/10/021.660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-128

Alignment Scores:
Pred. No.: 6,35e-36 Length: 411
Score: 549.00 Matches: 139
Percent Similarity: 50.12% Conservative: 64
Best Local Similarity: 34.32% Mismatches: 160
Query Match: 14.67% Indels: 42
DB: 14 Gaps: 10

US-09-645-078-1 (1-2043) x US-10-021-660-128 (1-411)
QY      143 AAAATGAAGCTCCTGCTGTTCTGCTTCC-----CAGATGGCCATCTTGGCTCTA 193
Db      6 LysAlaValLeuLeuLeuAlaLeuAlaSerIleAlaIleGlnTyrThrAlaIleArgThr 25
QY      194 TTCTCCCATGATACAGGCACAAATCAGTCCCTGCTTATCAGGCGACGCCCGCAGCGC 253

```

```

Db      26 PheThrAlaLysSerPheHisThrCysProGlyLeuAlaGluAlaGlyLeuAlaGluArg 45
QY      254 ATG-----CAGTGTCTGTT 268
Db      46 LeuCysGluGluSerProThrPheAlaTyrAsnLeuSerArgLysThrHisIleLeuIle 65
QY      269 CTGTCTTCTGCGCTCTGCGCTCTTCTTGTGGGGAGCTTTTGGGGAGCCAGAT 328
Db      66 LeuAlaThrThrArgSerGlySerPheValGlyGlnLeuPheAsnGlnHisLeuAsp 85
QY      329 GTTTTCTACCTGATGGAGCCGCTGCGACCTGTGGATGACC-----TTCAAG 376
Db      86 ValPheTyrLeuPheGluProLeuTyrHisValGlnAsnThrLeuIleProArgPheThr 105
QY      377 CAGACACCGGC-----TGGATGCTGCACATGGCTGTGGGATCTGATACGG 424
Db      106 GlnGlyLysSerProAlaAspArgValMetLeuGlyAlaSerArgAspLeuLeuArg 125
QY      425 GCGCTCTCTTGTGGACATGAGCGTCTTTGATGCTACATGGAACCTGTGCTCCCGGAGA 484
Db      126 SerLeuTyrAspCysAspLeuTyrPheLeuGlnAsnTyrIleLysProProValAsn 145
QY      485 CAGTCCAGC-----CTCTTTCAGTGGAGAACGCCGGCCCTGTGTCTTCTGCACCTGCC 538
Db      146 HisThrThrAspArgIlePheArgGlyAlaSerArgValLeuCysSerArgProVal 165
QY      539 TGTGACATCATC---CCACAAGATCAAATCATCCCGGGCTCACTGCAGGCTCTGTGTC 595
Db      166 CysAspProProGlyProAlaAspLeuValLeuGluGlyAspCysValArgLysCys 185
QY      596 ACTCAACAGCCCTTTCAGTGGTGGAGAGCGCTGCGCTCTTACAGCCAGCTGTGCTC 655
Db      186 GlyLeuLeuAsnLeuThrValAlaAlaGluAlaCysArgGluArgSerHisValAlaIle 205
QY      656 AAGGAGGTGCGCTTCTTCAACCTGCAGTCCCTACCGCTGCTGAAAGACCCCTCCCTC 715
Db      206 LysThrValArgValProGluValAsnAspLeuArgAlaLeuValGluAspProArgLeu 225
QY      716 AACCTGCATATCGTCACTCGTCCGGAGCCCGCGCGCTTTCCTGTCGCGAGAACGC 775
Db      226 AsnLeuLysValIleGlnLeuValArgAspProArgGlyIleLeuAlaSerArgSerGlu 245
QY      776 ACAAGGAGATCTCATGATGACAGTGCATTTGTGGGGCAGCATGACGACAAATC 835
Db      246 ThrPheArgAspThrTyrArgLeuTrpArgLeuTyrTyrGlyThrGlyArg----- 262
QY      836 AAGAAGGAGGACCAACCTACTATGTG-----ATGCAGGTCTATGCTGCAAGC 883
Db      263 -----LysProTyrAsnLeuAspValThrGlnLeuThrThrValCysGluAsp 278
QY      884 CAGCTGGAGATCTACAAGACCATCCAGTCTTGGCCAGGCCCTGCAGGAACGCTACCTG 943
Db      279 PheSerAsnSerValSerThrGlyLeuMetArgProProtrpLeuLysGlyLysTyrMet 298
QY      944 CTTGTGGCTATGAGGACCTGCGCTGAGCCCTGCGCCAGACTTCCCGAATGTATGAA 1003
Db      299 LeuValArgTyrGluAspLeuAlaArgAsnProMetLysLysThrGluGluIleTyrGly 318
QY      1004 TTCTGGGATGGAAATCTTGGCCCATCTTCAGACCTGGGTGCATACATCACCCGAGGC 1063
Db      319 PheLeuGlyIleProLeuAspSerHisValAlaArgTyrPheGlnAsnAsnThrArgGly 338
QY      1064 ---AAGGCATGGGTGACCGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCC 1120
Db      339 AspProThrLeuGlyLysHisLysTyrGlyThr---ValArgAsnSerAlaAlaThrAla 357
QY      1121 CAGCCTGGCCTGGTCTTTCGCTTATGAAAGCTTTCGACTTCAGAAAGCCTGTGCGC 1180
Db      358 GluLysTrpArgPheArgLeuSerTyrAspIleValAlaPheAlaGlnAsnAlaCysGln 377
QY      1181 GATGCCATGAATTTGCTGGCTACGCCACGTCAGATCTGTGACAAAGACAGAAACCTG 1240
Db      378 GlnValLeuAlaGlnLeuGlyTyrLysIleAlaAlaSerGluGluLeuLysAsnPro 397

```

QY 791 ATGATTGACAGTCGCATTGTGTGGGGCAGCATGAGCAAACTCAAGAGAGGACCA 850
Db 106 AlaArgAspAsnGluAlaala----- 112
QY 851 CCTACTATGTGATGCGAGTCTCTGCCAAAGCCAGCTGGAGATCTACAAGACCATCCAG 910
Db 113 -----ThrLeuLys 115
QY 911 TCCTTGGCCCAAGCCCTGCAGGAAAGCTACCTGTGTGGCTATGAGGACCTGGCTCGA 970
Db 116 Pro---ProProPheLeuArgGlyArgTyrArgLeuValArgPheGluAspLeuAlaArg 134
QY 971 GCCCTGTGGCCAGACTTCCCGAATGTATGAATTCGTGGGATTCGAATCTTGTGCCCAT 1030
Db 135 GluProLeuAlaGluLeuIleArgAlaLeuTyrAlaPheThrGlyLeuSerLeuThrProGln 154
QY 1031 CTTACAGACCTGGTGCATACATCACCCGAGCAAGGGCATGGGT 1075
Db 155 LeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGly 169

Search completed: May 6, 2004, 11:23:04
Job time : 109.5 secs

QY 1241 TTGCTGGATCTTCTG 1255
Db 398 SerValSerLeuVal 402
RESULT 15
US-09-927-602-6
; Sequence 6, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-6
Alignment Scores:
Pred. No.: 1,238-33 Length: 169
Score: 519.50 Matches: 113
Percent Similarity: 46.55% Conservative: 15
Best Local Similarity: 41.09% Mismatches: 34
Query Match: 13.88% Indels: 113
DB: 9 Gaps: 3
US-09-645-078-1 (1-2043) x US-09-927-602-6 (1-169)
QY 251 CGATGACAGTCGCTGGTCTGTCTCCCTGGCGCTCTGCTCTTCTTTGTGGGAGAGCTT 310
Db 8 ArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGlyLeuLeu 27
QY 311 TTGGGCGACACCCAGATGTTTCTTACCTGATGAGAGCCGCTGGCAGCTGTGGATGACC 370
Db 28 PheAsnGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrpThrThr 47
QY 371 TTCACGACGAGACCGCTGGATGCTGCACATGGCTGTGGGGATCTGATACGGGCGGTC 430
Db 47 ----- 47
QY 431 TTCTTGTGCGACATGAGGCTCTTTGATGCTTACATGGAACCTGGTCCCGGAGACAGTCC 490
Db 47 ----- 47
QY 491 AGCCTCTTTCAGTGGGAGAACAGCCGGGCCCTGTCTTCTGCACCTGCCTGTGACATATC 550
Db 47 ----- 47
QY 551 CCACAAGATGAATCATCCCCGGGCTCACTGCGAGGCTCCTGTGCGAGTCAACAGCCCTTT 610
Db 47 ----- 47
QY 611 GAGGTGGTGGAGAGGCTGCGCTCTACAGCCACGCTGTCTCAAGAGGTGCGCTTC 670
Db 48 -----LeuSerGluAlaCysArgSerTyrSerHisValValLeuLysGluValArgPhe 65
QY 671 TTCACCTCTGCTGCTCTACCGCTGTGAAAGACCCCTCCCTCAACCTGCATATCGTG 730
Db 66 PheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeuArgIleVal 85
QY 731 CACCTGGTCCGGGACCCCGGCGCTGTTCCTCCGAGAACGACAAAGGGAGATCTC 790
Db 86 HisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLysAlaLeu 105

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:57:23 ; Search time 24 Seconds
(without alignments)
8789.322 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 3742
Sequence: 1 gaattccattgtgttgga.....tgggattcctaaaaa 2043

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 389414 segs, 51625971 residues.

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODL=frame+n2p.model -DEV=xlpl
-Q=/cgn2_1/USPTO_spool_p/US09645078/runat_06052004_104600_21782/app_query.fasta_1.2183
-DB=Issued Patents AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-NODE=LOCAL -OUTEXT=pc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09645078.ecgn_1_1_27@runat_06052004_104600_21782 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Issued Patents AA:*
- 2: /cgn2_6/ptodata/2/iaa/5A.COMB.pap:*
- 3: /cgn2_6/ptodata/2/iaa/5B.COMB.pap:*
- 4: /cgn2_6/ptodata/2/iaa/6A.COMB.pap:*
- 5: /cgn2_6/ptodata/2/iaa/6B.COMB.pap:*
- 6: /cgn2_6/ptodata/2/iaa/6CTUS.COMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	54.5	386	3	US-09-045-284A-2
2	2038	54.5	386	4	US-09-190-911-1
3	1984	53.0	386	4	US-09-786-240-11
4	598.5	15.0	483	3	US-09-263-023-2
5	598.5	15.0	483	4	US-09-471-867-2
6	587.5	15.7	484	3	US-09-263-023-4
7	587.5	15.7	484	4	US-09-471-867-4
8	549	14.7	411	4	US-09-015-188-2
9	500.5	13.4	458	2	US-08-655-878-2
10	482	12.9	479	2	US-08-899-514-2
C 11	184	5.0	472	4	US-09-252-991A-31978
C 12	170	4.6	618	4	US-09-252-991A-23373

RESULT 1

US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6285192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

ALIGNMENTS

C 13	165.5	4.5	265	4	US-09-252-991A-32896	Sequence 32896, A
C 14	162.5	4.3	335	4	US-09-252-991A-23948	Sequence 23948, A
C 15	161.5	4.4	1034	4	US-09-252-991A-28921	Sequence 28921, A
C 16	160	4.3	232	4	US-09-252-991A-19777	Sequence 19777, A
C 17	159.5	4.3	421	4	US-09-252-991A-30742	Sequence 30742, A
C 18	157.5	4.3	504	4	US-09-252-991A-28242	Sequence 28242, A
C 19	157	4.2	341	4	US-09-252-991A-32424	Sequence 32424, A
C 20	157	4.2	465	4	US-09-252-991A-20576	Sequence 20576, A
C 21	155.5	4.2	301	4	US-09-252-991A-30700	Sequence 30700, A
C 22	155.5	4.2	492	4	US-08-468-995-12	Sequence 12, Appl
C 23	154.5	4.2	315	4	US-09-252-991A-20553	Sequence 20553, A
C 24	154	4.2	556	4	US-09-252-991A-22670	Sequence 22670, A
C 25	154	4.2	1089	4	US-09-252-991A-20334	Sequence 20334, A
C 26	153.5	4.1	369	4	US-09-252-991A-25394	Sequence 25394, A
C 27	153.5	4.1	448	4	US-09-252-991A-24066	Sequence 24066, A
C 28	152.5	4.1	215	4	US-09-252-991A-28157	Sequence 28157, A
C 29	152	4.1	582	4	US-09-252-991A-26182	Sequence 26182, A
C 30	151.5	4.1	170	4	US-09-252-991A-17086	Sequence 17086, A
C 31	151.5	4.1	250	4	US-09-252-991A-16684	Sequence 16684, A
C 32	151.5	4.0	663	4	US-09-252-991A-30843	Sequence 30843, A
C 33	151	4.0	226	4	US-09-252-991A-20432	Sequence 20432, A
C 34	150.5	4.0	297	4	US-09-252-991A-28842	Sequence 28842, A
C 35	150.5	4.1	529	4	US-09-252-991A-18245	Sequence 18245, A
C 36	150	4.0	256	4	US-09-252-991A-23974	Sequence 23974, A
C 37	150	4.1	267	4	US-09-252-991A-28620	Sequence 28620, A
C 38	150	4.1	591	4	US-09-252-991A-28760	Sequence 28760, A
C 39	149.5	4.0	681	4	US-09-252-991A-24567	Sequence 24567, A
C 40	149	4.0	319	4	US-09-252-991A-32635	Sequence 32635, A
C 41	149	4.0	442	4	US-09-252-991A-23285	Sequence 23285, A
C 42	148.5	4.0	432	4	US-09-252-991A-23622	Sequence 23622, A
C 43	147.5	3.9	375	4	US-09-252-991A-24278	Sequence 24278, A
C 44	147.5	4.0	433	4	US-09-252-991A-28695	Sequence 28695, A
C 45	147.5	4.0	491	4	US-09-252-991A-31758	Sequence 31758, A

Alignment Scores:

Pred. No.:	3,15e-191	Length:	386
Score:	2038.00	Matches:	386
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	54.46%	Indels:	0
DB:	3	Gaps:	0

US-09-645-078-1 (1-2043) x US-09-045-284A-2 (1-386)

Qy 128 ATGCTACTGCTTAAATAATGAAGTCCCTGCTTCTGTTTCCAGATGCCATCTTG 187
|||||
Db 1 MetLeuLeuProLysLysMetLysLeuLeuLeuLeuValSerGlnMetAlaIleLeu 20
|||||
Qy 188 GCTCTATTCTTCACATGTACGCCACACATCAGCTCCCTCTCTATGAGCAGCC 247
|||||


```
QY 728 GTGACCTGTCCGGAGCCCGGCGCTGTCCTCCGAGACGACAAAGGAGAT 787
Db 201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAsp 220
QY 788 CTCATGATTGACAGTCGCAATGTGATGGGCGAGCATGAGCAAAATCTCAAGAAAGGAGGAC 847
Db 221 LeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysGlyAsp 240
QY 848 CAACCTTACTATGTAGTCAGGTCATCTGCCAAAGCCAGTCGAGATCTCAAGACATC 907
Db 241 GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
QY 908 CAGTCCTTGGCCCAAGGCGCTGCAGCAACGCTACCTGCTTGTGGCTATGAGGACCTGGCT 967
Db 261 GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla 280
QY 968 CGAGCCCTGTGGCCGAGACTTCCCGAATGTATGAATCGTGGAATGGAATCTTGCC 1027
Db 281 ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro 300
QY 1028 CATCTTCAGACCTGGGTGCATACATCACCGAGGCAAGGCGCATGGGTGACCAAGCTTTC 1087
Db 301 HisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
QY 1088 CACAAATGCCAGGATGGCTTAAATGTCTCCAGGCTTGGCGCTGGTCTTTGCCCTAT 1147
Db 321 HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTyr 340
QY 1148 GAAAGGTTTCTGACATTCAGAAAGCTGTGGCGATGCCATGAATTTGCTGGGCTACCGC 1207
Db 341 GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrArg 360
QY 1208 CACGTCAGATCTGAACAGAACAGAGAAACCTGTGTGTCGATCTTCTGTACCTGGAAT 1267
Db 361 HisValArgSerGluGlnGluGlnArgAsnLeuLeuLeuAspLeuSerThrTrpThr 380
QY 1268 GTCCCTGAGCAATCCAC 1285
Db 381 ValProGluGlnIleHis 386

RESULT 3
US-09-786-240-11
; Sequence 11, Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11
Alignment Scores:
```

```
Pred. No.: 6.29e-186 Length: 386
Score: 1984.00 Matches: 383
Percent Similarity: 99.48% Conservative: 2
Best Local Similarity: 98.97% Mismatches: 1
Query Match: 53.02% Indels: 2
DB: 4 Gaps: 0
US-09-645-078-1 (1-2043) x US-09-786-240-11 (1-386)
QY 128 ATGCTACTGCTTAAAAAATGAAGCTCCTGCTGTTCTGTTTCCAGATGCCATCTTG 187
Db 1 MetLeuLeuProLysLysMetLysLeuLeuLeuPheLeuValSerGlnMetAlaIleLeu 20
QY 188 GCTCTATTCTTCCACATCTACAGCCACAAACATCAGCTCCCTGTCTATGAAGCCACAGCCC 247
Db 21 AlaLeuPhePheHisMetTyrSerHisAsnIleSerSerLeuSerMetLysAlaGlnPro 40
QY 248 GAGCCATGCAGCTGCTGTTCTCTCTTCCCTGGCGCTCTGGCTCTTCTTTTGTGGGCGAG 307
Db 41 GluArgMetHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGlyGln 60
QY 308 CTTTTTGGCGAGCACCCAGATGTTTCTACCTGATGGAGCCCGCTGGCACGTGTGATG 367
Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrpMet 80
QY 368 ACCTTCAAGCAGAGACACCGCTGCTGTCATGCTGCACATGGCTGCGGGATCTCATACGGGCC 427
Db 81 ThrPheLysGlnSerThrAlaTrpMetLeuHisMetAlaValArgAspLeuIleArgAla 100
QY 428 GTCTTCTTGTGGACATCAGAGGCTTTTCATGCTACATGGAACCTGTGTCAGTCAACAGCAG 487
Db 101 ValPheLeuCysAspMetSerValPheAspAlaTyrMetGluProGlyProArgArgGln 120
QY 489 TCCAGCCTCTTTCAGTGGGAGAACAGCGGCGCTGTTCTGTCACCTGCTGTGACATC 547
Db 121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuCysSerAlaProAlaCysAspIle 140
QY 548 ATCCACAGATGAATCATCTCCCGGGCTCAGTCAGGCTCCTGTGTCAGTCAACAGCAGCCC 607
Db 141 IleProGlnAspGlu-SerSerProGlyLeuThrAlaGlySerCysAlaValAsnSerPr 160
QY 608 TTTGAGGTG-GTGAGAGAGGCTGCGGCTCTCTACAGCCACGCTGTGCTCAAGGAGGTGCG 666
Db 160 oLeuLysLeuLeuGluLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
QY 667 CTTCTTCAACCTGTCAGTCCCTCTACCGCTGCTGAAAGACCCCTCCCTCAACCTGCATAT 726
Db 180 gPhePheAsnLeuGlnSerLeuTyrProLeuLysAspProSerLeuAsnLeuHisIle 200
QY 727 CGTGCACTGTGTCCGGGACCCCGGCGCTGTTCCGTTCCCGAGAACGACCAAAAGGGGAGA 786
Db 200 eValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAs 220
QY 787 TCTCATGATTGACATGTCGATTTGATGGGCGAGCATGAGCAAAATCTCAAGAGGAGGA 846
Db 220 pLeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAs 240
QY 847 CCAACCTTACTATGTAGTCAGGTCATCTGCCAAAGCCAGCTGGAGATCTCAAGACCAT 906
Db 240 pGlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
QY 907 CCAGTCTTGTGCCAAGGCGCTGCAGAAACGCTGCTGCTGTGCGCTATGAGGACCTGGC 966
Db 260 eGlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla 280
QY 967 TCGAGCCCTGTGGCCGAGACTTCCCGAATGTATGAATTTGTTGGGATTTGGAATCTTTC 1026
Db 280 aArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPr 300
QY 1027 CCATCTTCAGACCTGGGTGCATACATCACCGAGGCAAGGCGATGGGTGACCAAGCTTT 1086
Db 300 oHisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
```


Db 296 IleArgSerArgHisGlyLeuIleAArgGluSerLeuGlnValValArgSerArgAspPro 315
QY 815 -----GGCGACATGAGCAAACTCAGAGGAG 844
Db 316 ArgAlaHisArgMetPropheLeuGluAlaAAGlyHisLysLeuGlyAlaLysLysGlu 335
QY 845 -----GACCAACCTACTAT-----GTGATGAGTCTATCTGCCAAAGC 883
Db 336 GlyValGlyGlyProAlaAspTyrHisAlaLeuGlyAlaMetGluValIleCysAsnSer 355
QY 884 CAGCTGGAGATCTACAGACCATCCAGTCTTCCCAAGCCCTGCAGGAAGCTACCTG 943
Db 356 MetAlaLysThrLeuGlnThrAlaLeuGlnProProAspTrpLeuGlnGlyHisIstYrLeu 375
QY 944 CTGTGCGCTATGAGGACCTGGCTCGAGCCCTGTGCGCCAGACTTCCCGAATGTATGAA 1003
Db 376 ValValArgTyrGluAspLeuValGlyAspProValLysThrLeuArgArgValTyrAsp 395
QY 1004 TTGCTGGATTTGGAATCTTGGCCCATCTTCAGACCTGGGTGATACATCACCAGGC 1063
Db 396 PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly 415
QY 1064 AAGGCGATGGTGACCAAGCTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCCGAG 1123
Db 416 SerGlySerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn 435
QY 1124 GCTTGGCGCTGCTTTTGGCCCTATGAAGGTTTCTCGACTTCAGAAAGCCTGTGGGAT 1183
Db 436 AlaTrpArgThrAlaLeuThrPheGlnGlnIleLysGlnValGluGluPheCysTyrGln 455
QY 1184 GCCATGAATTTGCGGCTACCGCCAGCTGAGTCTCAACAAGAACAGAAACCTGTG 1243
Db 456 ProMetAlaValLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer 475
QY 1244 CTGGATCTTCTG 1255
Db 476 LysThrLeuLeu 479

RESULT 8

US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Alignment Scores:
Pred. No.: 4,52e-45 Length: 411
Score: 549.00 Matches: 139
Percent Similarity: 50.12% Conservative: 64
Best Local Similarity: 34.32% Mismatches: 160
Query Match: 14.67% Indels: 42
DB: 4 Gaps: 10

US-09-645-078-1 (1-2043) x US-09-015-188-2 (1-411)

QY 143 AAATGAAGCTCTCTGCTTTCTGTTCTCC-----CAGATGGCCATCTTGGCTCTA 193
Db 6 LysAlaValLeuLeuAlaLeuAlaSerIleAlaIleGlnTyrThrAlaIleArgThr 25
QY 194 TTCTTCCCATGTACAGCCACAAACATCAGTCCCTGTCTATGAAGGCACAGCCGAGCGC 253

Db 26 PheThrAlaLysSerPheHisThrCysProGlyLeuAlaGluAlaGlyLeuAlaGluArg 45
QY 254 ATG-----CAGTGTCTGTT 268
Db 46 LeuCysGluGluSerProThrPheAlaTyrAsnLeuSerArgLysThrHisIleLeuIle 65
QY 269 CTGCTTCTCTGCGCTCTGCTCTTCTTTGTGGGGCAGCTTTTGGGGCAGCACCAGAT 328
Db 66 LeuAlaThrThrArgSerGlySerPheValGlyGlnLeuPheAsnGlnHisLeuAsp 85
QY 329 GTTTTCTACCTGATGGAGCCGCTGCACGTGTGGATGACC-----TTCAAG 376
Db 86 ValPheTyrLeuPheGluProLeuTyrHisValGlnAsnThrLeuIleProArgPheThr 105
QY 377 CAGAGCACCGCC-----TGGATGCTGCACATGGCTGTGCGGATCTGATCAGG 424
Db 106 GlnGlyLysSerProAlaAspArgValMetLeuGlyAlaSerArgAspLeuLeuArg 125
QY 425 CCGCTCTTCTGCGGACATGAGCGTCTTTGATCGCTACATGGAACTGTGTCGCCGAGA 484
Db 126 SerLeuTyrAspCysAspLeuTyrPheLeuGlnAsnTyrIleLysProProValAsn 145
QY 485 CAGTCCAGC-----CTCTTTTCAGTGGGAGAACCGCGCCCTGTGTCTGTCACCTGCC 538
Db 146 HisThrThrAspArgIlePheArgArgGlyAlaSerArgValLeuCysSerArgProVal 165
QY 539 TGTGATCATC-----CCACAAGATGAATCATCCCGGGCTCACTGCAGGCTCTGTGC 595
Db 166 CysAspProProGlyProAlaAspLeuValLeuGluGluGlyAspCysValArgLysCys 185
QY 596 AGTCAACAGCCCTTTGAGTGTGTGAGGAAGCGCTCGCGCTCCTACAGCAGCTGTGCTC 655
Db 186 GlyLeuLeuAsnLeuThrValAlaAlaGluAlaCysArgGluArgSerHisValAlaIle 205
QY 656 AAGAGGAGTGGCTTCTTCACTGTCAGTCCCTCTACCGCTGTCTGAAGACCTCTCCTC 715
Db 206 LysThrValArgValProGluValAsnAspLeuArgAlaLeuValGluAspProArgLeu 225
QY 716 AACCTGCATATCGTCACCTGTGTCGGGACCGCGCGCTGTCTCCGCGAAGACGC 775
Db 226 AsnLeuLysValIleGlnLeuValArgAspProArgGlyIleLeuAlaSerArgSerGlu 245
QY 776 ACAAGGAGGATCTCATGATGACAGTGCAGTGTGATGGGGCAGCATGACCAAAATC 835
Db 246 ThrPheArgAspThrTyrArgLeuTrpArgLeuTrpTyrGlyThrGlyArg----- 262
QY 836 AAGAAGGAGGACCAACCTACTATG-----ATGCAGGTCTATCTGCCAAAGC 883
Db 263 -----LysProTyrAsnLeuAspValThrGlnLeuThrThrValCysGluAsp 278
QY 884 CAGCTGGAGATCTACAAGACCATCCAGTCTTGTGCCAAGCGCTGCAGGAACGCTACCTG 943
Db 279 PheSerAsnSerValSerThrGlyLeuMetArgProProTrpLeuLysGlyLysTyrMet 298
QY 944 CTGTGCGCTATGAGGACCTGCTCGAGCCCTGTGCGCCAGACTTCCCGAATGTATGAA 1003
Db 299 LeuValArgTyrGluAspLeuAlaArgAsnProMetLysThrGluGluIleTyrGly 318
QY 1004 TTCTGCGGATTTGGAATCTTGTGCCCATCTTACAGCTGGGTGTCATACATCACCAGGC 1063
Db 319 PheLeuGlyIleProLeuAspSerHisValAlaAlaArgTyrIleGlnAsnThrArgGly 338
QY 1064 ----AAGGCGATGGTGCACCGCTTTCACACAAATGCAGGATGCGCTTAATGCTCC 1120
Db 339 AspProThrLeuGlyLysHisLysTyrGlyThr-----ValArgAsnSerAlaAlaThrAla 357
QY 1121 CAGGCTTGGCGTGGTCTTTGCGCTTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGCG 1180
Db 358 GluLysTrpArgPheArgLeuSerTyrAspIleValAlaPheAlaGlnAsnAlaCysGln 377
QY 1181 GATGCCCATGAATTTGCTGGGCTACCGCCACGCTCAGTGTGAACAGACAGAGAACCTG 1240

Db 378 GlnValLeuAlaGlnLeuGlyTyrLysIleAlaAlaSerGluGluGluLeuLysAsnPro 397

Qy 1241 TTGCTGATCTTCG 1255

Db 398 SerValSerLeuVal 402

RESULT 9

US-08-655-878-2

; Sequence 2, Application US/08655878

; Patent No. 5927713

; GENERAL INFORMATION:

; APPLICANT: FUKUTA, MASAKAZU

; APPLICANT: HABUCHI, OSAMI

; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE:

; STREET:

; CITY:

; STATE:

; COUNTRY:

; ZIP:

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/655,878

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME:

; REGISTRATION NUMBER:

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE:

; TELEFAX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 458

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-655-878-2

Alignment Scores:

Pred. No.: 2,750-40 Length: 458

Score: 500.50 Matches: 120

Percent Similarity: 54.02% Conservative: 75

Best Local Similarity: 33.24% Mismatches: 133

Query Match: 13.38% Indels: 33

DB: 2 Gaps: 13

US-09-645-078-1 (1-2043) x US-08-655-878-2 (1-458)

Qy 227 CTGCTATGAGGACACCCGAGCGATGACGTGCTGCTGCTTCTCTGCGGCTCT 286

Db 104 LeuGlyIleAlaAlaProGluProArgHisValLeuLeuMetAlaThrThrArgThr 123

Qy 287 GCCTCTCTTTTGTGGGCGAGCTTTTGGGCGAGCACCCAGATGTTTCTACTCATGGAG 346

Db 124 GlySerSerPheValGlyGluPheAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 143

Qy 347 CCGCGCTGGCAGTG-----TGGATGACCTTCAAGCAGACACCGCGCTGGATGCTGAC 400

Db 144 ProLeuTphHisIleGluArgThrValThrPheGluProGlyGlyAlaAsnAlaValGly 163

Qy 401 ATGGCTGTG-----CGGATCTGATACGGCGCTCTTCTGTGCGATGAGGTCTTT 454

Db 154 SerAlaLeuValTyrArgAspValLeuGlnGlnLeuLeuLeuLeuLeuLeuLeuLeu 183

Qy 455 GATGCTACATGGAACCTGTCCTCCGAGAGAG-----TCCAGCCTCTTTCAGTGGAG 508

Db 184 GluSerPheIleSerProAlaProGluHisLeuThrAlaAlaLeuPheArgGly 203

Qy 509 AACAGCGCGCGCTGCTGTCGACCTGCTGTCACATCATCCACAA---GATGAAATC 565

Db 204 SerSerHisSerLeuCysGluGluProValCys-----ThrProSerLeuLysVal 221

Qy 566 ATCCCGCGGCTCATCTGAGG---CTCCTGTGAGTCAACAGCCCTTTGAGGTGGTGGAG 622

Db 222 PheGluLysTyrHisCysLysAsnArgCysGlyProLeuAsnIleThrLeuAlaAla 241

Qy 623 AAGGCTGCGCTCTCTCAGCCACGTGTGCTCAAGGAGGTGCGCTTCTTCAACCTGAG 682

Db 242 GluAlaCysArgArgLysGlnHisMetAlaLeuLysThrValArgIleArgGlnLeuGlu 261

Qy 683 TCCTCTACCCGCTGCTGAAAGCCCTCCCTCAACCTGCATATGCTGCACCTGCTGCGG 742

Db 262 PheLeuGlnProLeuAlaGluAspProArgLeuAspLeuArgIleIleGlnLeuValArg 281

Qy 743 GACCCCGCGCGCTGCTCCGTTCCGAGAACGCAAGGGAGATCTCATGATTGACAGT 802

Db 282 AspProArgAlaValLeuValSerArg-----MetVal----- 292

Qy 803 CGCATTGTGATGGGCGCAGCATGACGAAAAAACACTCAAG----- 838

Db 293 ---AlaPheSerGlyLysTyrGluSerTrpLysTyrTrpAlaAlaGluGlyGluAlaPro 311

Qy 839 ---AAGGAGGACCAACCTACTATGTGATGAGTCACTGCCAAGC---CAGTGGAG 892

Db 312 LeuGlnGluAspGlu---ValGlnArgLeuArgLysAsnCysGluSerIleArgLeuSer 330

Qy 893 ATCTACAGACCATCCAGTCTTCCCAAGCGCTTCGAGGAAACGCTACCTCTTGTGCGC 952

Db 331 AlaGluLeuGlyLeuArgGln---ProArgTrpLeuArgGlyArgTyrMetLeuValArg 349

Qy 953 TATGAGACCTGCTGAGCCCTGTCGCCAGACTTCCCGAATGTATGATTCGTGGGA 1012

Db 350 TyrGluAspValAlaArgAlaProLeuArgLysAlaLeuGluMetTyrArgPheAlaGly 369

Qy 1013 TTGGAATCTTCCCGCATCTTCCAGACCTGGGTGCATACATCATCCGAGCAAGGCGCATG 1072

Db 370 IleHisProThrProGlnValGluGluTrpIleArgAlaAsnThrGlnAlaPro---Gln 388

Qy 1073 GGTGACACGCTTCCACACAAATCCAGGAGTCCCTTAATGTCTCCAGGCTTGGCGC 1132

Db 389 AspSerAsnGlyIleTyrSerThrGlnLysAsnSerSerGluGlnPheGluLysTrpArg 408

Qy 1133 TGGTCTTTGCCCTATGAAAAGGTTTCTGACTTCAGAAAGCTGTGGCGCATGCAATGAAT 1192

Db 409 PheSerIleProPheLysLeuAlaGlnValGlnAspAlaCysGluProAlaMetArg 428

Qy 1193 TTGCTGGGTACCGCCAGCTGATCTGAACAGAACACAGAAACCTCTTGTGATCTT 1252

Db 429 LeuPheGlyTyrLysLeuAlaSerSerAlaGlnGluLeuThrAsnArgSerLeuSerLeu 448

Qy 1253 CTG 1255

Db 449 Leu 449

RESULT 10

US-08-899-514-2

; Sequence 2, Application US/08899514

; Patent No. 5910581

; GENERAL INFORMATION:

; APPLICANT: HABUCHI, OSAMI

; APPLICANT: FUKUTA, MASAKAZU

; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN

; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING

; TITLE OF INVENTION: FOR THE POLYPEPTIDE

; NUMBER OF SEQUENCES: 9

US-09-645-078-1 (1-2043) x US-09-252-991A-31978 (1-472)

US-05-0435-078-1 (1-2043) X US-05-252-591A-31978 (1-472)

QY	1142	GCAAGACCAAGCGCAAGCCTGGAGACATTAAAGGCAATCCCTGGCATTTGTGTGGAAG	1083
Db	55	AlaProArgGlyAlaGlyProGlyProGlyAlaArgPro	68
QY	1082	CGTGGTCACCCATGCCCTTGCTCGGTGATGTTATCACCCAGGCTGAAGATGGGCA	1023
Db	69	ArgGlyAspArgAlaAspArgAlaGlySerAlaThrSerProArgAlaPro	86
QY	1022	AGAATCCAAATCCACGAATTCATACATTCGGGAAGTCTGGG	981
Db	87	ProGlyProSerProHisProGlyValProGlyAlaThrAlaAspAlaThr	104
QY	980	CCACAGGGCTCGAGCCAGGTCCTCATACGCGCACAGCAGGTAGCGTT	933
Db	105	ProGlyAlaAlaProGlySerValThrValArgAlaHisArgGluSerAlaSerAlaGly	124
QY	932	CCTCAGGGCT	900
Db	125	ThrAlaArgProArgArgHisTTPAlaAlaThrAlaValArgSerAlaArgHisAla	144
QY	899	TGTAGATCTCCAGCTGGCTTGGCAGATGACCTGCATCACATAGTAGGTTGGTCCCTCT	840
Db	145	AlaAlaArgProAlaAlaAlaLeuArgGlnProAlaProHisSerGlyGlySerProPro	164
QY	839	TCTTGAGTTTTGTCTATGCTGCCCATCACAATCGCACTGCTCAATCATGAGATCTCCCT	780
Db	165	ThrSerAlaLeuArgHis	176
QY	779	TTGTGCGTCTCCGGAACGGA	738
Db	177	GlyAlaAlaAlaGlyAsnGlyArgArgAlaAlaArgProGlyArgProGlyGluSerPro	196
QY	737	CCAGGTGCAGTATGCAAGT	684
Db	197	LeuAlaGlyGlnGluArgAlaGlyValAlaProAlaValGlnTTPAlaProProAla	216
QY	683	ACTCAGGTTGAAGAGCGGCACTTCCTTGACACCACTGCTGTAGAGCGGGCAGG	627
Db	217	ThrAlaGly	233
QY	626	CTTCTCCACCACTCAAGGGCTGTTGACTGCACAGGA	588
Db	234	AlaSerArgLysArgProProProAlaProProArg	246
QY	587	GCCTGCAGTAGCCGGGGATGATTTCATCTTGGGATGATGTCACAGCGGTCCAG	528
Db	247	ArgSerThrProGlyAlaProAlaAlaHisProGlyGluProProAlaAlaProProLe	252
QY	527	AACACAGGGCCGGCTGTTCCCACTGAAAGAGGCTGGAATGCTCCGGGACCAAGT	468
Db	253	ArgSerThrProGlyAlaProAlaAlaHisProGlyGluProProAlaAlaProProLe	272
QY	467	CAATGTAGCAT	444
Db	273	ArgProGlyProValArgArgValProProAlaThrArgProAlaSerAspArgAla	292
QY	443	TGTGCGC	405
Db	293	LeuArgThrProAlaGlySerValArgArgProProAlaLysArgGlyGlnSer	312
QY	404	CAATGTGACGA	384
Db	313	ProAlaProAlaAlaAlaValProAlaGlyAlaAlaProAlaLeuArgArgArgAla	332
QY	383	TGCTCTGTCTGAAGTCTACACAGCTGCCAGGCGGCTCCATCAGG	327
Db	333	AlaThrAlaAlaAlaAspSerArgLeuAlaArgAlaProAlaGlyArgArgHis	352
QY	326	CTGGGTGTCGCCAAAAAGTGCCTCCCAACAAAGAGCCAGAGCC	279
Db	353	AlaAlaProAlaSerProAlaArgProValProAlaAlaLeuAlaAlaAlaGlyAsn	372

RESULT 12

US-09-252-991A-23373
: Sequence 23373, Application US/09252991A

; Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELAT

; TITLE OF INVENTION: AERUGIN

FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1999-03-23

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142

```

; NUMBER OF SEQ
; SEQ TD NO 23373

```

```

; SEQ ID NO 233
; LENGTH: 618

```

```

; LENGTH: 010
; TYPE: PRT

```

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23373

Alignment Scores:

Pred. No.:	8.82e-08	Length:	618
Score:	170.00	Matches:	133
Percent Similarity:	32.88%	Conservative:	35
Best Local Similarity:	26.03%	Mismatches:	191
Query Match:	4.59%	Indels:	152
DB:	4	Gaps:	29

US-09-645-078-1 (1-2043) x US-09-252-991A-23373 (1-618)

Qy	1585	CTCCAGGCCACAAAGTCTCTGAAGTCTGCCACACAGGAAGATCAAGAAAAGNAGAGA	1526
Db	158	ProProthrHis-----ArgArgA:garg	165
Qy	1525	GCAATACCCGTGTTCACTGCTGGG-----	1502
		:::	
Db	166	AspLeuProAlaIatnRgLyArgHisTyrProGlyArgLeuGlnLeuGlyProGlyGly	185
Qy	1501	-----ATGTCGTCTCACATAGGTTCCCGAGTCTGTTTTCTAGACACAAAGCATGGACA	1448
		:::	
Db	186	GlnPheArg-ProProGlnLeuArgLeuProProLeuLeuSerArg-----GlyYH	202
Qy	1447	CAAAAGTCCTTCTGCTTTGACGACGGTGGGACACAACCTCACCTCAGTGTGATACCCACAG	1388
Db	202	sAlaGlyProProGlyv--AlaLeuLeuArqProGlyvHis---Hi sleuAspGlnAlaAr	220


```
QY 858 ATGTGATCGAGGTGATCTGCCAAGCCAGCTGGAGATCTACAAGACCATCCAGTCCTTGC 917
   ...
Db 298 lile-----
QY 918 CCAAGGCCCTCAGGAACGCTACCTGCTGTGCGCTATGAGACCTGCTCGAGCCCTG 977
   |||||
Db 299 --ArgTTPCysGlyThrThrIleCys-----TTPArgCysTTP--ArgProPro 313
   |||||
QY 978 TGGCCAGACCTCCGGAATGATGAAATTCGTTG-----GATTGGAATTCCTGCCCATC 1031
   |||||
Db 314 TTPProAla-----AlaTTPProGlySerTTPArgThrCys----- 325
   |||||
QY 1032 TTCAGACCTGGGTGCATACATCACCC 1058
   |||||
Db 326 -----TTPSerCysArgThrSerPro 332
   |||||

RESULT 15
US-09-252-991A-28921
; Sequence 28921, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28921
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28921

Alignment Scores:
Pred. No.: 7.9e-07 Length: 1034
Score: 161.50 Matches: 122
Percent Similarity: 31.94% Conservative: 38
Best Local Similarity: 24.35% Mismatches: 194
Query Match: 4.36% Indels: 147
Ds: 4 Gaps: 24

US-09-645-078-1 (1-2043) x US-09-252-991A-28921 (1-1034)
QY 1588 AGGCTCCAGGCCACAAAGTCTCTGAAGTCTGCCAGACAGGAAGATCAAGAAAGAAGA 1529
   |||||
Db 32 ArgProAlaSerHisAlaGlnArgAlaAspArgAspArgLeu---ArgArgGln 50
   |||||
QY 1528 AGAGCAATACCTGTTCTCACTGGTGGATGCTGCTCAC-----ATAAGTT 1481
   |||||
Db 51 ArgAla--TyrProGlyAlaAlaValGlyArgProAlaHisArgAlaGlnProAlaArg 70
   |||||
QY 1480 CCCAGTCTGTTTCTAGACACAAGCATGGACACAAAAGTCTCTGCTTCCAGCACGTGT 1421
   |||||
Db 70 GProGlnThrPheAsp-----HisAlaArgAlaArgProAlaArg----- 83
   |||||
QY 1420 GGACACAACCTACACTAGTGTGATACCCACAGACATGTAGTTAAGGCTCAGAGATGTAG 1361
   |||||
Db 84 -----ArgProArgAlaProGlyProArgProAlaG 95
   |||||
QY 1360 GCAAGGCTCAGAAGCATTCAGAGAAGTGACTGAGGCTGACACAGGTGGCGAGCAAGGCC 1301
   |||||
Db 95 lYThrGlyArgProGlyProArgProProAlaAlaAspGlnArgAlaGlnArgThrA 115
   |||||
QY 1300 TTCTCAACCTCTTAGTGGATTGCTCAGGACAGCTCCAGGTAGACAGAAATCCAGCAA 1241
   |||||
Db 115 rg-----AlaProGlyHisProGlnArgProArgAlaPro----- 126
   |||||
QY 1240 CAGGTTTCTCTGTTCTGATCTGACGTGGCGGTAGCCAGCAAAATTCATGGCATC 1181
   |||||
```

```
Db 127 -----GlyGlyAlaProArgAlaArgA 136
   |||||
QY 1180 GCCACAGGCTTCTGAAGTCGAGAAACCTTTTCATAGGCGAAGACAGCCGCAAGCCTG 1121
   |||||
Db 136 rGHisAlaGluGlyArgSerGlyAlaAlaGlyGlyTTPGlnLeuProGlnGlyArgProG 156
   |||||
QY 1120 GGAGACATTAAGGCGATCCCTGGCATTGTTGTGGAAAGCGTGGTCACCATGCCCTGCC 1061
   |||||
Db 156 lYArgHis-----ProProAlaGlnAlaArgArgGlyArgAlaAlaProLeu----- 171
   |||||
QY 1060 TCGGTGATGTTATGCACCCAGGTCTGAAGTGGGCAAGATTCATC----- 1011
   |||||
Db 172 --\-----ArgArgGlyArgArgLeuProAlaAlaG 185
   |||||
QY 1010 -----CCACGAATTCATACATTCGGGAAGT 986
   |||||
Db 185 luArgProArgProArgGlnProArgProAlaGlyProArgArgHisAlaGlyGlyArgS 205
   |||||
QY 985 CT-----GGGCCACAGGGGCTC---GAGCCAGGTCTCTATACGCGACAG 944
   |||||
Db 205 erArgGluArgProAlaAlaProAlaAlaLeuGlyArgProGlyLeuLeuAlaAlaProA 225
   |||||
QY 943 CAGGTAGCTTCTCTGACGGGCTTGGGCAAGACTGGATGCTTGTAGATCTCCAGCTG 884
   |||||
Db 225 la---AlaAlaProAlaGlyArgLeuCysArgThrAlaArgLeuAlaValLeuProAlaA 244
   |||||
QY 883 GCTTTGGCAGATGACCTGCATCATCATAGTAGGTTGGTCTCTCTCTTTAGTTTGTCT 824
   |||||
Db 244 laValAspAlaGluProAlaAlaThrGlyHisGlyLeuProValArgArgProLeuAlaA 264
   |||||
QY 823 ATGCTGCCCCATCACAAATGCGACTGCTCATCATGAGATCTCCCTTTGCGTCTCTCGGA 764
   |||||
Db 264 laSerGlyProAlaArg-----ProAlaProAla----- 273
   |||||
QY 763 ACGGAACACGGCCCGGGGT----- 744
   |||||
Db 274 --AlaThrGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 293
   |||||
QY 743 -----CCGGACACAGGTGCGATATGAGGTGAGGAGGGGTCTTTGAGCAG 695
   |||||
Db 293 roValAlaLeuProGlyProArgValArgArgGlyProGlyLeuArgProArgArgP 313
   |||||
QY 694 CGGTTAGAGGGACTGCGAGTTCAAGAGCGCACCTCTTGGACACACCGTGGCTGTAGGA 635
   |||||
Db 313 roGlyArg-----ArgSerLeuGlnPro-----T 321
   |||||
QY 634 GGGCAGGCTTCTCCACCA-----CCTCAAGGGCTGTTGACTGCA 593
   |||||
Db 321 rPdlnArgProGlyProProGlyArgThrGlySerArgArgArgArgLeuArgThrSerP 341
   |||||
QY 592 CAGGACCTGCGAGTACGCGGGGGATGATTTCATCTTGGGATGATGTCACAGCGCAGG 533
   |||||
Db 341 roGlyAlaProAlaAlaThrGly-----ArgArgA 351
   |||||
QY 532 TGCAAGAACACAGGGCCCGGC-----TGTCTCCCACTGAA---AGAGCTCGACTG 485
   |||||
Db 351 laThrGlnGlnGlyProGlyArgGlyThrAlaAlaProThrProGlyThrGlyAlaA 371
   |||||
QY 484 TCTCCGGGACACAGGTTCCATGTAGCATCAAGAGCTCATGTGCGCAAGAGAGCGC 425
   |||||
Db 371 laAlaGly-----ArgArgGlnGlnGlyThr-----AlaP 381
   |||||
QY 424 CCTATTCAGATCCCGCACAGCCATGTGCAGCATCCAGCGGTCTGCTGCTGAAGGTCTAT 365
   |||||
Db 381 roGlySerGluProAlaGluProAlaAlaArgGlnProAlaAlaAlaArgArgG 401
   |||||
QY 364 CCACAGCTCCAGCGGGGCTCCATCAGTAGAAACATCTGGGTGCTGCCAAAAGCTG 305
   |||||
Db 401 lYAspArgArgCysAlaGluGlyProGlyArg-----GlyAlaLeuAspGlnProA 418
   |||||
QY 304 CCCCAACAAAGAGAGCCAGAGCGCGCAGGAACACAGACAGACCGTGCATCGCTCGG 246
   |||||
```

us-09-645-078-1.ra1

Thu May 6 11:30:32 2004

Db 418 laGlyArgArgArgGluAspArgArgSerThrArgProSerAlaAlaAlaGln 437

Search completed: May 6, 2004, 11:08:13
Job time : 42 secs

JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; transmembrane protein

Query Match 29.8%; Score 606.5; DB 2; Length 486;
Best Local Similarity 37.0%; Pred. No. 3.9e-44;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERMHVLSSWRSGSFVGGQFGQHPDVFLMEPAWHVMMTFKOSTAWMLHMAVRDLIRA 100
DB 100 EKQHYVHATWRTGSGFLGELFNQNPVEFLYEPVHWQKLYPGDAVSLQGAARDMLSA 176
QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SSLFQWNSRALCSAPACDIIPDEIIIPRAHCRLL- 155
DB 177 LYRCDLSVQLYSPAGSGRNLTLTGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDDRCK 235
QY 156 -CSQOPFEVVEKACRSYSHVVLKEVRFENLOSFLPKDPSLNLIHIVLVRDPRVAFRSR 214
DB 236 KCPORLARFECECKRYTLVKGVRVDFVAVLAEPLLDLPALDKVILHVRDPRVAFSSR 295
QY 215 ERTKGLMIDSRIVM-----GQHEOKLKE-----DQPY-----VMQVICS 252
DB 296 IRSRGLRESLQVRSRDPRAHRMFFLEAAGHKLGAKEGVGGPADYHALGAEVICNS 355
QY 253 QLEIKYTIOSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELPHLOTHVNIIRG 312
DB 356 MAKTUQTALQPDWQGHYLVVRYEDLVGBPVKTRRVIDFVGLLVSEMEQFALNMTSG 415
QY 313 KGMGDHAFHTNARDALNVSAQWWSLPYKVSRLQKACGDAMNLLGYRHVRSEQQRNLL 372
DB 416 SGSSSKFFVVSARNATQAANAWRTALTQQIKQVEEFCYQPMVAVLGYRVNSPEEVKOLS 475
QY 373 LDLL 376
DB 476 KTL 479

RESULT 4
A57397
Chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C:Species: Gallus gallus (Chicken)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
C:Accession: A57397
R;Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi,
J. Biol. Chem. 270, 18575-18580, 1995
A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransf
A:Reference number: A57397; MUID:95355490; PMID:7629189
A:Accession: A57397
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-458 <FUK>
A:Cross-references: GB:D49915; NID:G971262; PIDN:BAA08655.1; PID:G971263
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match 24.6%; Score 500.5; DB 2; Length 458;
Best Local Similarity 33.2%; Pred. No. 4.6e-35;
Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

QY 34 LSMKAQPERMHLVLSWSRSGSFVGGQFGQHPDVFLMEPAWHV--WMTFKQSTAWMLH 91
DB 104 LGTAAPRPRHULLMATTGSGFVGEFNQGNIFLYLEPLWHIERTVTFEPGGANAVG 163
QY 92 MAV--RDLIARVFLCDMSVFDAYMEPGPRRQ--SSLFQWNSRALCSAPACDIIPQ-DEI 146
DB 164 SALVYRDVLQQLLDCDLYLESFISPAPEHLTAALFRGSGSHLCEEPVC--TPSLKKV 221
QY 147 IPRAHCR-LICSQOPFEVVEKACRSYSHVVLKEVRFENLOSFLPKDPSLNLIHIVHLVR 205
DB 222 FEKHCNRCRGGPLNITLAEACRRQKHVALKTVIRIQLEFLOPLAEDPLDLRIIQLVR 281
QY 206 DPRVAFRSRERTKGLMIDSRIVMVGQHEQKLK-----KEDQPYVMQVICS-QUE 255
DB 282 DPRVAVLSR-----MV---APSGYESKWKMAEAGEAPLOEDE-VQRLGNCSIRLS 330
QY 256 IYKTIOSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELPHLOTHVNIIRGKM 315
DB 331 AEGLGKQ-PRWLRGRTMLVRYEDVAPRPAKALEMTFRAGIHPTFQVEEWIRANTQAP-Q 388
QY 316 GDHAFHTNARDALNVSAQWWSLPYKVSRLQKACGDAMNLLGYRHVRSEQQRNLLDL 375

JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; transmembrane protein

Query Match 29.8%; Score 606.5; DB 2; Length 486;
Best Local Similarity 37.0%; Pred. No. 3.9e-44;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERMHVLSSWRSGSFVGGQFGQHPDVFLMEPAWHVMMTFKOSTAWMLHMAVRDLIRA 100
DB 100 EKQHYVHATWRTGSGFLGELFNQNPVEFLYEPVHWQKLYPGDAVSLQGAARDMLRS 159
QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SSLFQWNSRALCSAPACDIIPDEIIIPRAHCRLL- 146
DB 160 LFRCDFSVRLYAPGPDPAARAPDTANLTALFRWTRTKVICSPPPLCPGAPARAEEVGL 219
QY 147 IPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVRFENLOSFLPKDPSLNLIHIVHLVRD 206
DB 220 VEDTACERSCPPVALEAECKYPVVVKDRLDGLVPLLRDPLGLNKKVQLFRD 279
QY 207 PRAVFRSERTKGLMIDSRIVM-----GQHEOKLKEEDQPY 244
DB 280 PRAVHNSRLKSRQGLRESIQVLRTRQGRDFRHLVLAHGVGARPGGSRALPAAPRADF 339
QY 245 ----VMQVICSQOLBIYKTIOSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELP 300
DB 340 FLTGALVCEAWLDDLPAFGAPAWLRRVRLRYEDLVQPRALRLRLRFSGLRALA 399
QY 301 HLTQWVHNITRKGWG-DHAFHTNARDALNVSAQWWSLPYKVSRLQKACGDAMNLLGY 359
DB 400 ALDAPALANMGAAYGADRPFLSARDAREAVHAWRELSREQVRQVEACAPAMRLIAY 459
QY 360 RHRVSEQE 367
DB 460 --PRSGEE 465

RESULT 3
JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C:Accession: JE0261
R;Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Pan, Q.W.; Mitsuoka,
T.
J. Biochem. 124, 670-678, 1998
A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
A:Reference number: JE0261; MUID:98391845; PMID:9722682
A:Accession: JE0261
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DDBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfate from 3'-phosphoadenosine
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match 28.9%; Score 588.5; DB 2; Length 484;
Best Local Similarity 36.0%; Pred. No. 1.4e-42;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

Db 389 DNGIYSTOKNSSEFEKWRFSIPPKLAQVVDQADPAPNRLFGYKLAASAGELNRSLSL 448
QY 376 L 376
Db 449 L 449

RESULT 5
E95934
Probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sinc
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95934
R:Pinan, T.M.; Weidner, S.; Wong, K.; Ruhmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95934
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:gl5140626; GSPDB:GN00167
A:Experimental source: Strain 1021, megaplasmid pSymB
R:Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb21237
A:Genome: plasmid

Query Match 8.2%; Score 168; DB 2; Length 307;
Best Local Similarity 22.7%; Pred. No. 9.2e-07;
Matches 83; Conservative 50; Mismatches 131; Indels 102; Gaps 16;

QY 36 MKAQPRMHLVLSNRSGSFVQGLFGQHPDVF---YLMPEAPHWMTFKQSTAWMLHM 92
Db 1 MFSQVRI-AYTAGYGRSGTILDALGQHAAVGVANGSITSLTRVWRHNEYCA----- 53
QY 93 AVRDILRAVFLCDMSVFDAYMEPPGRQSSLFQWNSRALCSAPACDIPODEIIPRAHC 152
Db 54 -----CGNAIRDCSFWSVRR-----EWSDG-----QDPGLMEBYC 84
QY 153 RLICSQPPEVVEKACR-----SYSHVVLKEVRPFN-LQS----- 186
Db 85 AL---QQKEGLSMVTRLLSGMLGKQFSLYLHTKRLPSAMQSCSGRQIVDSKLPGR 141
QY 187 LYPLLKPSLNHLVHLVDPRAVFRS-----RERTKGLMIDSRIVMGQHEQKKEKED 240
Db 142 AMAVAQIPGIDMRVHLVDRGRGVAWSLLKGVYERDAKSG-----LQKEI 185
QY 241 QPYVMQVTCQSQLEYTKTQSLRAL-QERYLLVRYEDLAPAPVAQTSRMVYFVGLFL 299
Db 186 KPKSVFRTALRWSM-VNLAVEYLSRKLGSKEKVMRYVEDFASDPVAVMQQIGTFLFELD-L 243
QY 300 PHLTQVWHNITRGKMG-----DHAFHTNARDALNVSQAWRSFLPYEKSRLQKACGDA 353
Db 244 SQVGTSLN-GEAMGPGHGVAGNRLRMVASIALNKDETWTMPARQVQSFQLGGWM 300
QY 354 MNLGY 359
Db 301 LRRYG 306

RESULT 6
G70729
hypothetical protein RV2267c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70729
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squires, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70729
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-388 <COL>
A:Cross-references: GB:Z77163; GB:AL123456; NID:93261610; PIDN:CAB00968.1; PID:e255072;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV2267c

Query Match 5.7%; Score 116; DB 2; Length 389;
Best Local Similarity 19.3%; Pred. No. 0.037;
Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

QY 45 VLVLSNRSGSFVQGLF---GQH--PDVFLMEPAHWMTFKQSTAWMLHVAVDLIR 99
Db 83 IFIVGHWRTGTTLLHLLVVDNRHTGTGYECLAP--HHPLL-----TEW----- 125
QY 100 AVFLCDMSVFDAYMEPPGRQSSLFQWNSRALCSAPACDIIPQ-DEII-----PRAHC 152
Db 126 -----FAPVVE-----FLVSKHRANDMDLSLHHPQEDFVCMQGLPSPYL 167
QY 153 RLICSQP-----FEVFKACRSYSHVVLKEVRFNLSLY-----PLLKDPSLNLIH 200
Db 168 TIAPFNRPPQYEEYLDLEQVAPRELE--IWKRTLRFVQCVYFRRTKTVILKNPHTSFRI 225
QY 201 -----VHLVDRPRAVFRSERTKGLMIDSRIVMGQHEQKKEQDPYY--VM 246
Db 226 KVLEVPQAKFHIVRDPVYV-----PSTHLKALYRIHGLQOPTDGLD 273
QY 247 QVTCQSQLEYTKTQSLPKALQ-ERYLLVRYEDLAPAPVAQTSRMVYFVGLF-----FLPH 301
Db 274 DKVSTVVDLYRLKDLREGELVDPTRFVELYEDLIGDPGQLRLRYQHILGLGDFECLYLR 333
QY 302 LQTVWHNITRGKMGDHA-PHTNARDALNVSQAWRSFLPYEKSRLQKACGDMNLGY- 359
Db 334 LRQ-----LADHADYKTN-----YQLTVQRAIVDEHWGEIIDRYGYD 373
QY 360 RH 361
Db 374 RH 375

RESULT 7
I40485
Surfactin synthetase component I - Bacillus subtilis
N:Alternate names: competence protein srfAA; surfactin production protein srfAA; surfaci
C:Species: Bacillus subtilis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000
C:Accession: I40485; B63718; S35517; A37323; S46367; A43705; S34985
R:Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perrego, M.; Venema, G.; van Sin
Mol. Microbiol. 8, 821-831, 1993
A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
A:Reference number: I40485; MUID:93360813; PMID:8355609
A:Accession: I40485
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3586 <RES>
A:Cross-references: EMBL:X70356; NID:9396480; PIDN:CAA49816.1; PID:G396481
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galled
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.P.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A.; Mauch, P.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel
 V.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadai, Y.; Sato, T.; Scanlon,
 A.; Schleib, S.; Schroeder, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Serr
 akuchii, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A;Authors: Yoshikawa, H.F.; Zumeitein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A;Reference number: A69580; MUID:98044033; PMID:9384377
 A;Accession: B59718
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-3588 <KUN>
 A;Cross-references: GB:Z91905; GB:AL009126; NID:G2632457; PIDN:CAB12142.1; PID:G2632634
 A;Experimental source: strain 168
 R;Fuma, S.; Fujishima, Y.; Corbell, N.; D'Souza, C.; Nakano, M.M.; Zuber, P.; Yamane, K.
 Nucleic Acids Res. 21, 93-97, 1993
 A;Title: Nucleotide sequence of 5' portion of *srfA* that contains the region required for
 A;Reference number: S35517; MUID:93181186; PMID:841623
 A;Accession: S35517
 A;Status: nucleic acid sequence not shown; significant sequence differences
 A;Molecule type: DNA
 A;Cross-references: EMBL:D13262; NID:G216345; PID:G216346
 A;Experimental source: strain 168 trpC2
 A;Note: protein sequence not complete; the nucleotide sequence was submitted to the EMBL
 R;Nakano, M.M.; Magnuson, R.; Myers, A.; Curry, J.; Grossman, A.D.; Zuber, P.
 J. Bacteriol. 173, 1770-1778, 1991
 A;Title: *srfA* is an operon required for surfactin production, competence development, an
 A;Reference number: A37323; MUID:91154134; PMID:1847909
 A;Accession: A37323
 A;Status: nucleic acid sequence not shown; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-46, EV, 49-145, I, 147-150, L, 152, AN, 155-280, T, 282-307, SF, 310-384
 A;Cross-references: GB:M5939; NID:G14366
 R;Nakano, M.M.; Xia, L.; Zuber, P.
 J. Bacteriol. 173, 5487-5493, 1991
 A;Title: Transcription initiation region of the *srfA* operon, which is controlled by the
 A;Reference number: A43705; MUID:91358326; PMID:1715856
 A;Contents: annotation
 R;Pabret, C.; Quentin, Y.; Guiseppe, A.; Busuttil, J.; Haeich, J.; Denizot, F.
 submitted to the EMBL Data Library, March 1993
 A;Reference number: S46967
 A;Accession: S46967
 A;Molecule type: DNA
 A;Residues: 3249-3271, A, 3273-3316, R, 3318-3451, Y, 3453-3483, DE, 3486-3487, DAGL, 34
 A;Cross-references: EMBL:X72672; NID:G516358; PIDN:CAA51222.1; PID:G516359
 A;Experimental source: strain 168 trpC2
 A;Comment: this enzyme is one of several in the multienzyme complex that synthesizes the
 nce.
 C;Genetics:
 A;Gene: *srfAA*; *srfAL*
 A;Note: *srfAA* is the first gene of the *srfA* operon
 C;Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein h
 C;Keywords: antibiotic biosynthesis; carrier protein; phosphopantetheine; phosphoprotein
 F;507-956/Domain: acetate-CoA ligase homology <ACL1>
 F;574-1042/Domain: acyl carrier protein homology <ACP1>
 F;1043-1488/Domain: repeat <RPT1>
 F;1549-1993/Domain: acetate-CoA ligase homology <ACL2>
 F;2011-2079/Domain: acyl carrier protein homology <ACP2>
 F;2080-2527/Domain: repeat <RPT2>
 F;2589-3025/Domain: acetate-CoA ligase homology <ACL3>
 F;3042-3109/Domain: acyl carrier protein homology <ACP3>
 F;1006,2043,3074/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

990 VEKAGIFNPFETGGHSLKAMTLLTKIHK-----ETGIEIPQQLFPHPTITALAE 1040
 74 PAWHVMTFTKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPGPRQSSLSFQWNSRALC 133
 1041 EADH-----RESKAF-----AVIEFAEKQEHYPLHMSSEHIS 1073
 134 SAPA-----CDIIPQDBII-----PRAHCRLLCS 157
 1074 SASSRMSRAIHASSNSSEGRFSYKAGARISGINPTFVITSTFVLENSTPROKIH-VCV 1132
 158 QQPFVEVKACRSYSHVVLKEVRFNLOSPLKDPSSL-----NLIHV-----HLVRDPR 208
 1133 DFNIEWIERGGRSDAIIASVFTFDLAKA-PLFRIGLLGLEENRHMLLFDMHLLISGV 1191
 209 ANFRSERTKGLMD--SRIVMGHEQKLKEDQPYVMQVICOQSLKIEYKTIQSLPKA 266
 1192 SI-----GIMLEELARIYKGEQLDLRLQYKDYAVWQ--SRQAAEGYKQDAYWKE 1240
 267 L--QERYLLVRYEDLARAPV 284
 1241 VFAGELFVLQLLSDYPRPPV 1260
 RESULT 8
 T13288
 mei-41 protein - fruit fly (*Drosophila melanogaster*)
 C;Species: *Drosophila melanogaster*
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C;Accession: T13288
 R;Hari, K.L.; Santerre, A.; Sekelsky, J.J.; McKim, K.S.; Boyd, J.B.; Hawley, R.S.
 Cell 82, 815-821, 1995
 A;Title: The mei-41 gene of *D. melanogaster* is a structural and functional homolog of th
 A;Reference number: Z11072; MUID:95401271; PMID:7671309
 A;Accession: T13288
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-2354 <HAR>
 A;Cross-references: EMBL:U34925; NID:G998351; PID:G998353; PIDN:AAC46881.1
 C;Genetics:
 A;Gene: mei-41
 A;Cross-references: FlyBase:FBgn0004367
 A;Introns: 650/3; 748/3; 2313/3
 C;Function:
 A;Description: involved in cell cycle checkpoint and meiotic recombination
 Query Match 4.5%; Score 92.5; DB 2; Length 2354;
 Best Local Similarity 20.5%; Pred. No. 42;
 Matches 85; Conservative 52; Mismatches 147; Indels 131; Gaps 20;
 28 SHNISS--LSMKAQPERMHVLVLSSWRSGSFVGLFGQHPDVFYLMPEAMHVMVMTFKQ 84
 1058 NYNFAFPQHLPLNLSDDFAVLALTSLCRGYQF--QNTKHYVDSFSL----- 1102
 85 STAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRQSSLSFQWNSRALCSPACDIIPOD 144
 1103 -----AIQETLA---ICGIS-----PKECKYQLWQ-----SLPA---RMKQ 1133
 145 EIIPRAHCRLLCSQCPFEVVEKAC--RSYSHVVLKEVRFF-----NLOS 186
 1134 LMPEMLHSCYTCVHRSTCLOQPLFGSHYSHNYEAWFLMASRLIDYLPSSGKRLLSS 1193
 187 LYPLKX-----PSLNLIHV-----HLVRDPAVRSRRTKRGDLMIDSRIV 228
 1194 YKFCIKRDSNMLSTFFYPYIILLHALLCTTEQRNHIQEEFMAVLQANES-----SSSV 1246
 229 MGQHEQKLKEDQPYVMQVICOQSLKIEYKTIQSLPKALQERYLLVRYEDLARAPVAQTS 288
 1247 RGQELGAIKEN-----AFKQTESRYAAGI-KPLASTLVSDRKEDSRVP----- 1291
 289 RMVEFVGLLEFLPHLOTW-----VHNITRGKMGDHAFTNAR---DALNVQAMWWSLP 339
 1292 RLAGKLCAELDLFLQWLREWQIRHGRSTG-GKPPETIDSNVRKIHFEFLNL-----IP 1343

Query Match 4.7%; Score 95; DB 2; Length 3588;
 Best Local Similarity 19.7%; Pred. No. 45;
 Matches 63; Conservative 48; Mismatches 111; Indels 98; Gaps 13;

14 VSQMAILAFHMYSHNISLSMKAQPERMHVLVLSSWRSGSFVGLFGQHPDVFYLM 73

QY 340 YEKVSRLOKACGDAMNLLGYRHVRSQ-----EQNLLDLLSTWTVPEQI 385
DB 1344 KLLVSRASVNCGEVARALSYLESYLEGEDSKORLLSQFTFLVVEVYGLRDPDSV 1398
RESULT 9
T16350
hypothetical protein F42G9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C:Accession: T16350
R:Taich, A.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid F42G9.
A:Reference number: Z18498
A:Accession: T16350
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-359 <TAI>
A:Cross-references: EMBL:U00051; NID:G1216305; PID:G1216308; PIDN:AAA91354.1; CESP:F42G9
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F42G9.8
A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8
Query Match 4.5%; Score 92; DB 2; Length 359;
Best Local Similarity 19.5%; Pred. No. 3.8; Mismatches 114; Indels 98; Gaps 15;
Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;
QY 52 RGSFVQQLFCQHPDV----FYLMEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMS 107
DB 95 RSGTLMRAILDAPDVRCGGTMLPFLTWQAGWRND-WYNSGI-----TQE 143
QY 108 VDAYMEPGPRQSSFLQWNSRALCSAPACDIIPQ-DEIIPRAHCRLLCSQOQPFVEVEX 166
DB 144 VDDAV-----SAFTEIVAKHSELAPR-----LCKDP----- 172
QY 167 ACRSYSHVYLKEVRFNFLQSLYPLKDPSLNLHIVHLVRDPRAVRSRRTK----GDLM 222
DB 173 ----YTALWLPTR-----RLYP-----NAKFLIMRIDARAVVHSMIERKVPVAGYNT 216
QY 223 IDSRVWGQHEQKLKEDQPYVYMQVICOQ-----LEIYKTIQSLPKALQERYLLVRYED 278
DB 217 SDIEISFVQWQELRK-----MTFCNNAPGQCIVY-----YER 251
QY 279 LARAPVAQTSRMVYFVGLFELPHLQTVVHNITRGKMGDHAPH-TNARDALNVQAMRWS 337
DB 252 LIQKPAEETLRITNFDLPFSQMLRHQDLIGDEVLDNQPSASQVKNSTNTRKALTSWF 311
QY 338 LPY-EKVSLQACGDAMNLLGY 359
DB 312 DCFSEETLRKLDVAPFLGILGY 334
RESULT 10
JE0196
hydroxysteroid sulfotransferase (EC 2.8.-.-) - mouse
C:Species: Mus musculus (house mouse)
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 18-Jun-1999
C:Accession: JE0196
R:Sakakibara, Y.; Yanagisawa, K.; Takami, Y.; Nakayama, T.; Suiko, M.; Liu, M.C.
Biochem. Biophys. Res. Commun. 247, 681-686, 1998
A:Title: Molecular cloning, expression, and functional characterization of novel mouse
A:Reference number: JE0196; MUID:98321187; PMID:9647753
A:Accession: JE0196
A:Molecule type: mRNA
A:Residues: 1-338 <SAK>
A:Cross-references: GB:AF026072; NID:G2570897; PIDN:AAC69918.1; PID:G2570898
C:Comment: This enzyme catalyzes the transfer of a sulfonate group from the active sulfa
p.
C:Superfamily: alcohol sulfotransferase
C:Keywords: transferase

Query Match 4.5%; Score 91.5; DB 2; Length 338;
Best Local Similarity 22.2%; Pred. No. 3.8; Mismatches 32; Indels 71; Gaps 11;
Matches 52; Conservative 32; Mismatches 32; Indels 71; Gaps 11;
QY 67 DVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMBPGPRRQSSLFQW 126
DB 59 DIFIVTYP-----KSGTNWMB-----IVCLILKD-----GDPMSIRSPWQ- 96
QY 127 ENSRALCSAPACD-IIPQDEIIPRAHCRLLCSQOQPFVEVEXKACRSYSHVYLKEVRFENLQ 185
DB 97 -----RAPWCETIISAFNVLDPRSPRIMSHLPIELFTKA-----FF--- 133
QY 186 SLVPLKDPSLNLHIVHLVRDPRAVRSRRTKGDLMIDSRVWGQHEQKLKEDQPYVY 245
DB 134 -----SSKAKVIYVGNRPDVVS-----LYYSKIA-GQ-----LKDPGTPOQF 172
QY 246 MQVICQSQLEIYKTIQSLPKAL-----QERYLLVRYEDLARAPVAQTSRMVYFVG 295
DB 173 LQNFLKGEVQFGWFHDKIRMGIRMQENFLFITYEELQDDLGRGVSQVRCIFLG 226
RESULT 11
S11737
resistance protein Mx3, interferon-regulated - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Mar-2000
C:Accession: S11737
R:Meier, E.; Kunz, G.; Haller, O.; Arnheiter, H.
submitted to the EMBL Data Library, April 1990
A:Description: Unexpected activity of rat Mx proteins against a Rhabdovirus.
A:Reference number: S11735
A:Accession: S11737
A:Molecule type: mRNA
A:Residues: 1-659 <EMB>
A:Cross-references: EMBL:X52713; NID:G56724; PIDN:CAA36937.1; PID:G56725
C:Superfamily: dynamin-related protein VPS1

Query Match 4.5%; Score 91; DB 2; Length 659;
Best Local Similarity 20.6%; Pred. No. 10; Mismatches 50; Indels 96; Gaps 16;
Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;
QY 62 FGQHPDVFVYLMPEPAWHVMTFKQSTAWMLHMAVRDLI-----RAVFLCDMSVFDAYME--- 114
DB 301 FKHPQFRALLEDG-----KATVPCLAERLTMELSHICKSLPLENQIKESHQSTSE 353
QY 115 -----PGPRRQSSFLQWNSRALCSAPACDIIPQDEIIPRAHCRLLCS-QQPF--- 161
DB 354 ELQKYGADIPEDENEXTFLIEKINAF-NQDITAVEGEEIVREKECRLLTKLKEFFLW 412
QY 162 -EVVEKACRSYSHVYLKEVRFENLQ-----SLYPLLKDPKPSLNLHI 200
DB 413 SEEIERNFQKSDALYKEVYTFEMQYRGRELPGFVNYKTFENIIRQIKTLSEPAEM--- 470
QY 201 VHLVRD-PRAVPR-----SRRTKGDLMIDSRVWGQHEQKLKEDQPYVYMQV 248
DB 471 LHKVTEIVRAAFTVSEKNFSEPFNLHRTTKSKLEDIRL---EQETEAKAIRLHFQMQ 527
QY 249 ICOSQLEIYKTIQSLPKALQERYLLVRYEDL-----ARAPVAQTSRMVYFVG 296
DB 528 IIVCQDQIYR-----KALQK-----VREBEEERKKGKRSQAQSPNLTSSMD----- 572
QY 297 EFLPHLQTV---VHN 308
DB 573 EITQHLNAYRQEAHN 587
RESULT 12
B64477
replication factor C homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: B64477

```

Db      73  KFFKARNDNFVCSYKPTGTTWLKALTPA-----IANRSKEDFVSTNPLLRNPEHF-- 123
QY      127  ENSRALCASPADI-----IPQDIIIPRAHCRLLCSQPPE-----VWEKACRSYSHVVK 177
Db      124  -----VPYIEIDPFPPFSDVLKDEGNTLFSTHPIYDLLPESVVKSGCK----- 167
QY      178  EVRFENLQSLYPLLKDPSLNLHIVHLVRDPRPRAVFRS-----RERTKGDLMLDSRIYVMGQ 231
Db      168  -----IVYIWRDPKOTFVSMWTFFAHKERSQQQPVVSTEEAFDK 205
QY      232  HEQKLKEDQYYVMQVICSQLEIVKTTQSLPKALOERYLLVRVEDLARAPVAQTSLRMV 291
Db      206  YQGLSAYG-PY-----LDHVGVMKAYQANP-----DOILEKLETWRADPLPYVKLA 254
QY      292  FEVLGLEF-----LPHLTQVWYHNITRGKMGDHAFTNARDALNYSQA--- 333
Db      255  EPMGYGFTKEEGNVEKVVKLCSPET-LKQLEANKGEKDR-----EDRPAVYANSAYFR 309
QY      334  -----WWSLPIYEKVSR 346
Db      310  KGKVGDWQNYLTPEMWARI 328

RESULT 14
T07120
N:Contains: cytochrome P450 CP7 - soybean
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
C:Accession: T07120
R:Schopfer, C.R.; Ebel, J.
Mol. Gen. Genet. 258, 315-322, 1998
A:Title: Identification of elicitor-induced cytochrome P450s of soybean (Glycine
A:Reference number: Z15931; MUID:98311068; PMID:9648734

```

RESULT 14

T07120

probable cytochrome P450 CP7 - soybean
N;Contains: oxidoreductase (EC 1.-.-.)
C;Species: Glycine max (soybean)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
C;Accession: T07120
R;Schopfer, C.R.; Ebel, J.
Mol. Gen. Genet. 258, 315-322, 1998
A;Title: Identification of elicitor-induced cytochrome P450s of soybean (Glycine max L. cv. Williams 82) and their role in the biosynthesis of isoflavonoid phytoalexins
A;Reference number: Z15931; MUID:98311068; PMID:9648734
A;Accession: T07120
A;Status: unaligned;trans. translated from GR/EMBL/DBD

A; Residues: 1-504 <SCH>
A; Cross-references: EMBL:Y10493; NID:Q3334666; PIDN:CAA71517.1; PID:G3334667
A; Experimental source: cultivar Hazoev 63, strain Marr
C; Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C; Keywords: Chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase
F; 303-466/Domain: cytochrome P450 homology <P45>
F; 444/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 4.4%; Score 89.5; DB 2; Length 504;
Best Local Similarity 19.9%; Pred. No. 9.7;
Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16

Qy 68 VYFLMEPAHWMTKQSTAWMLHMAVDLIRAVFLCDMSVPDAYMEPGPROSSLFQWE 127
Db 14 VFLLHL-WLV-KTYQKSSHL-----PPGWR----- 39
Qy 128 NGRALCSAPACDIIIPQDSIIIPRAHCLLCQQPFVVEKACRSYS---HVVLKEVRFNVL 184
Db 40 -----LP---IIGNLHQLAAASLPDQALQKLVKRYGFLMHLGLCEISTLVV 83*
Qy 185 QSLYPLLKDPSLNLHIVHLVDRPAVFRSRRTKGLMIDSRVTWGHSQKLKKEQDPQY 244
Db 84 SS--PKQANEMMKTHDVHFVRPQ-----LLAPQFVYVYQATDIAFAPYGDYR 129
Qy 245 VMGVICQSQLETIKYTIQSLPALKQE--RYLLVRVEDIARAPVACTSRMVEYVGL----- 296

DB		130	QIRKITLELLSAKVQSFHSIRQDENKKLIQSITHSSAGSPIDLSGKLFSLLIGTTVSRAA	189
QY		297	-----EPL-----PHLOTWVHNITRGKMGDGAHHTNA----	324
DB		190	FGKENDQDEFFSLVRKAITWTGGFEVDNPFSLKP-LHLITLTQAKVERV-HQRADKIL	247
QY		325	RDALNSQAARWSLPYEKVSLOKACGDAMNLGYRHVRSEQORNLLDLL	376
DB		248	EDILRKHM-----EKRTRYKEGNG-----SEAQGEDLVDTL	279

Search completed: May 6, 2004, 10:51:57
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:46:52 ; Search time 17 Seconds
(without alignments)
1182.299 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKKMLLLFLVSQMAIL.....EQRNLLDLLSTWTVPQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	500.5	24.6	458	C6ST_CHICK	Q92179 gallus gall
2	116	5.7	388	YM67_MYCTU	Q50695 mycobacteri
3	104.5	5.1	419	NOEE_RHISN	P55472 rhizobium s
4	95	4.7	358	SRF1_BACSU	P27206 bacillus su
5	92	4.5	359	TPSB_CAEEL	Q20351 caenorhabdi
6	91.5	4.5	5430	MACF_HUMAN	Q9up13 homo sapien
7	91.5	4.5	5938	MACF_HUMAN	Q96pk2 homo sapien
8	91	4.5	294	239F_HUMAN	Q15777 homo sapien
9	91	4.5	639	MX3_FAT	P18590 rattus norv
10	89.5	4.4	504	C7D8_SOYBN	O81974 glycine max
11	89.5	4.4	4273	PKSM_BACSU	P40872 bacillus su
12	89	4.4	320	F4ST_FLACH	P52837 flaveria ch
13	89	4.4	524	MPPA_RAT	P20069 rattus norv
14	89	4.4	920	CAPP_STRPY	Q9a0u7 streptococ
15	89	4.4	1968	RRPO_PMR	P17965 porato viru
16	88	4.3	312	F3ST_FLABI	P52835 flaveria bi
17	88	4.3	2209	KNTC_HUMAN	P50748 homo sapien
18	87.5	4.3	380	TPSC_CAEEL	O77081 caenorhabdi
19	87	4.3	940	CHRD_BRARE	O57472 brachydanio
20	86.5	4.2	748	MUTA_MOUSE	P16332 mus musculu
21	86.5	4.2	831	DPOF_TRETH	P30113 thermus the
22	86	4.2	823	SCH9_YEAST	P11792 saccharomyc
23	85.5	4.2	559	BAL2_ARATH	Q9asr4 arabidopsis
24	85	4.2	632	GLCE_CAEEL	P46555 caenorhabdi
25	85	4.2	932	CAPP_STRP3	Q8k873 streptococ
26	85	4.2	1237	YDY2_SCHPO	O13683 schizosacch
27	85	4.2	2688	ZEPI_MOUSE	Q03172 mus musculu
28	84.5	4.1	646	SVT_LACLA	Q9ced2 lactococcus
29	84.5	4.1	780	CUL5_MOUSE	Q9d5v5 mus musculu
30	84.5	4.1	2493	YBA4_YEAST	P35194 saccharomyc
31	83.5	4.1	803	GLND_CANAL	Q94065 candida alb
32	83.5	4.1	864	GLND_PASMU	Q9cuh1 pasteurella
33	83	4.1	222	NOGG_XENLA	P49011 xenopus lae

ALIGNMENTS

RESULT 1

ID	C6ST_CHICK	STANDARD;	PRT;	458 AA.
AC	Q92179;1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).			
OS	Gallus Gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
FN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.			
RC	STRAIN=White leghorn; TISSUE=Embryonic chondrocytes;			
RX	MEDLINE=95355490; PubMed=7629189;			
RA	Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,			
RA	Shinomura T., Habuchi O.;			
RT	"Molecular cloning and expression of chick chondrocyte chondroitin 6-			
RT	sulfotransferase.";			
RL	J. Biol. Chem. 270:18575-18580(1995).			
CC	- - FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN			
CC	SULFATE.			
CC	- - CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin =			
CC	adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.			
CC	- - SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By			
CC	similarity).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; D49915; BAA08655.1; --			
DR	PIR; A57397; A57397.			
DR	InterPro; IPR000863; Sulfotransferase.			
DR	Pfam; PF00685; Sulfotransfer; 1.			
FT	Transferase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.			
FT	DOMAIN 1 23			
FT	CYTOPLASMIC (POTENTIAL).			
FT	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)			
FT	(POTENTIAL).			
FT	TRANSMEM 24 37			
FT	DOMAIN 38 458			
FT	LUMENAL, CATALYTIC (POTENTIAL).			
FT	CARBOHYD 62 62			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 73 73			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 95 95			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 236 236			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 399 399			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	CARBOHYD 443 443			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
SQ	SEQUENCE 458 AA; 52253 MW; C9A3B7D0A5086F0C CRC64;			
Query Match	24.6%;	Score	500.5;	DB 1;
Best Local Similarity	33.2%;	Pred. No.	1.7e-34;	Length 458;

Matches	120;	Conservative	75;	Mismatches	133;	Indels	33;	Gaps	13;
Qy	34	LSMAQPERHYLVLSWSSGSSVFGOLFGQHPDVFVLMPEAHV--WMTFKQSTAWMLH	91						
Db	104	LGIAAPRPRHLLMATRTGSSVFGFFNQGNIFLFLPWHIERTVTFEGGANVG	163						
Qy	92	MAV--RDLIRAVFLCDMSVFDAYMEPRRQ--SSLFQWNSRALCSAPACDIPO-DEI	146						
Db	164	SALVYRDVQLLCLDYLLESFISPAPEEHLTAALFRGGSSHSLCEEPVC--TPSLKKV	221						
Qy	147	IPRAHOR-LLCSQOPFEWEKACSYSHVLFKEVRFENFENLQSLVPLKLPDLNLIHVLVR	205						
Db	222	PERKHCNRRGPNITLAEACRKHMAKTVIRIQLEFLOPLADPLDLRIQLVR	281						
Qy	206	DPRAVFRSRETKGLMIDSRVNGHQEKLK-----KEDQPYVNVQVCS-QLE	255						
Db	282	DPRAVLSR-----NV---APSGKYESMKMAEAGEAPLQDE-VQRLRGNCESIRLS	330						
Qy	256	LYKTIQSLPKALQRYLLVRYEDLARAPVAQTSRMYEFVGLFELPLHLOTWVHNITRGKM	315						
Db	331	AELGLRQ-PRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHTPFPQVEWIRANTQAP-Q	388						
Qy	316	GDHAFHTNARDALNVSOAWRNSLPYKVSRLQKACGDAMNLLGYRHRVRSQEQORNLLDL	375						
Db	389	DSNGIYSTQKNSSEQFEKWRISIPFKLAQVVDACEPAMRLFGYKLASSAQELNRSLSL	448						
Qy	376	L 376							
Db	449	L 449							

RESULT 2

Ym67 MYCTU	STANDARD;	PRT;	388 AA.
AC	Q50695;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Hypothetical protein RV2267C/MT2329/Wb2290C.		
GN	RV2267C OR MT2329 OR MTCY339.43 OR MB2290C.		
OS	Mycobacterium tuberculosis, and		
OS	Mycobacterium bovis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773, 1765;		
RN	[1]_TaxID=1773, 1765;		
RP	SEQUENCE FROM N.A.		
RC	SPECIES=M.tuberculosis; STRAIN=H37Rv;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence."		
RL	Nature 393:537-544 (1998).		
RN	[2]		

SEQUENCE FROM N.A.

RP	SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RC	MEDLINE=22206494; PubMed=12218036;
RX	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA	Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA	Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT	laboratory strains."
RL	J. Bacteriol. 184:5479-5490 (2002).
RN	[3]

RP	SEQUENCE FROM N.A.
RC	SPECIES=M.bovis; STRAIN=AF2122/97;
RX	MEDLINE=22703107; PubMed=12788972;
RA	Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA	Prior M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA	Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA	"The complete genome sequence of Mycobacterium bovis."
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	
DR	EMBL; 271163; CAB00968.1; ..
DR	EMBL; AE007076; AAK46611.1; ..
DR	EMBL; BX248342; CAD97151.1; ..
DR	FIR; G70729; G70729.
DR	TIGR; MT2329; ..
DR	TubercuList; RV2267C; ..
DR	InterPro; IPR000863; Sulfotransferase.
DR	Pfam; PF00685; Sulfotransfer; 1.
KN	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 388 AA; 46062 MW; 5DED0263275A9B24 CRC64;
Qy	45 VLVLSSWSSGSSVFGOLF---GQH--PDVFLMEFAMVWMTFKQSTAWMLHMAVRDLIR 99
Db	83 IFVGHWRIGTLLHLLVDDRHGTGTCYELAP--HFFLL---TEW----- 125
Qy	100 AVFLCDMSVFDAYMEPRRQSSLFQWNSRALCSAPACDIPO-DEI-----PRAHC 152
Db	126 -----FAPYVE-----FLVSKHRAMNDLSLHHPQDEDFVWCQGFSPVL 167
Qy	153 RLCSOOP-----PEVVEKACSYSHVLFKEVRFENFENLQSLY-----PLKDPGLNLIH 200
Db	168 TIAPFNRPPQYBEYLDLEQVAPRELE--IWKTLFRFVQVYFRRKTVILKNPTHSFRI 225
Qy	201 -----VHLVRDPRAVFRSRETKGLMIDSRVNGHQEKLKKEEQPY--VM 246
Db	226 KVLEVPQAKFTHIVRDPVYVY-----PSTHLKALYRIHGLQQTDFDGLD 273
Qy	247 QVTCQSLQLEIKYTIQSLPKALQ-EXLLVRYEDLARAPVAQTSRMYEFVGLS-----FLPH 301
Db	274 DKVYSTVDLYRKLDGRELVDPTRFYELKRYEDLDIGDEGQLRRLYQHLGLGDFCYLPR 333
Qy	302 LOTWVHNITRGKMGMDHA-FHTNARDALNVSOAWRNSLPYKVSRLQKACGDAMNLLGY- 359
Db	334 LRQY-----LADHADYKINS-----YQLTVEQRAIVDEHGWGIIDRYGYD 373
Qy	360 RH 361
Db	374 RH 375
Qy	
Db	

Query Match 5.7%; Score 116; DB 1; Length 388;
 Best Local Similarity 19.3%; Pred. No. 0.025;
 Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

RESULT 3

NOEE_RHISN	STANDARD;	PRT;	419 AA.
ID	NOEE_RHISN		
AC	P55472; P72326;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	Nodulation protein noe (EC 2.8.2.-).		
GN	NOE OR Y4HS.		
OS	Rhizobium sp. (strain NGR234).		
OG	Plasmid sym pNGR234a.		

SEQUENCE OF 1-460 FROM N.A.
MEDLINE=91154134; PubMed=1847909;
Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
Zuber P.;
"srfa" is an operon required for surfactin production, competence
development, and efficient sporulation in *Bacillus subtilis*.";
J. Bacteriol. 173:1770-1778(1991).
[6]
RN SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=168;
RA MEDLINE=95219080; PubMed=7704255;
Fujishima Y., Yamane K.;
"A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
of srfaA of the *Bacillus subtilis* chromosome.";
Microbiology 141:277-279(1995).
[7]
RN SEQUENCE OF 1-38 FROM N.A.
RC MEDLINE=91358326; PubMed=1715856;
Nakano M.M., Xia L., Zuber P.;
"Transcription initiation region of the srfa operon, which is
controlled by the comp-comA signal transduction system in *Bacillus*
subtilis.";
J. Bacteriol. 173:5487-5493(1991).
CC - FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC - COFACTOR: Contains 3 covalently bound phosphopantetheines.
CC - PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
CC - SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC - SIMILARITY: Contains 3 acyl carrier domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D13262; BAA02522.1; -
EMBL; X70356; CAA49816.1; -
EMBL; D50453; BAA08982.1; -
EMBL; Z99105; CAB12142.1; -
EMBL; M59939; AAA22815.1; -
EMBL; D30762; BAA21034.1; -
EMBL; M64702; AAA22816.1; -
PIR; I40485; I40485.
HSP; P14687; 1AMU.
Subtilast; BG10168; srfaA.
InterPro; IPR00873; AMP-bind.
InterPro; IPR001242; Condensatn.
InterPro; IPR006163; Pp bind.
InterPro; IPR006162; Ppantne S.
Pfam; PF00501; AMP-binding; 3.
Pfam; PF00658; Condensation; 4.
Pfam; PF00550; Pp-binding; 3.
PRINTS; PRO0154; AMPBINDING.
PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
PROSITE; PS00455; AMP BINDING; 3.
PROSITE; PS00075; ACP DOMAIN; 3.
DR Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW Multifunctional enzyme; Repeat; Complete proteome.
FT REPEAT ? 1047 DOMAIN 1 (GLU-ACTIVATING).
FT REPEAT ? 2084 DOMAIN 2 (LEU-ACTIVATING).
FT REPEAT ? 3115 DOMAIN 3 (D-LEU-ACTIVATING).
FT DOMAIN 976 1043 ACYL CARRIER (ACP) 1.
FT DOMAIN 2013 2080 ACYL CARRIER (ACP) 2.
FT DOMAIN 3044 3110 ACYL CARRIER (ACP) 3.
FT BINDING 1006 1006 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 2043 2043 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 3074 3074 PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 146 146 M -> I (IN REF. 1 AND 5).

FT	CONFLICT	151	151	V -> L (IN REF. 1 AND 5).
FT	CONFLICT	165	165	G -> A (IN REF. 1).
FT	CONFLICT	281	281	Q -> T (IN REF. 1 AND 5).
FT	CONFLICT	460	460	D -> T (IN REF. 5).
FT	CONFLICT	540	540	A -> P (IN REF. 1).
FT	CONFLICT	562	562	Y -> I (IN REF. 1).
FT	CONFLICT	639	640	PT -> GS (IN REF. 1).
FT	CONFLICT	644	644	P -> R (IN REF. 1).
FT	CONFLICT	647	649	FLP -> LLA (IN REF. 1).
FT	CONFLICT	1026	1026	Q -> L (IN REF. 1).
FT	CONFLICT	1065	1114	HWSHSEISSASSMRESAISASSNSSEGFYSKAGARIS INPTPRVIE -> SLAQRTYIVSQFEDAGVYNNPAAAIL EGPLDIQKLERAFQGLIRRHESLR (IN REF. 1).
FT	CONFLICT	1130	1131	VC -> DS (IN REF. 1).
FT	CONFLICT	1182	1162	A -> R (IN REF. 1).
FT	CONFLICT	1456	1456	L -> V (IN REF. 1).
FT	CONFLICT	1848	1848	T -> S (IN REF. 1).
FT	CONFLICT	1892	1892	T -> D (IN REF. 1).
FT	CONFLICT	1909	1909	A -> R (IN REF. 1).
FT	CONFLICT	1971	1978	PAVFQMD -> LRCLSKWT (IN REF. 1).
FT	CONFLICT	2050	2050	P -> L (IN REF. 1).
FT	CONFLICT	2216	2216	Q -> H (IN REF. 1).
FT	CONFLICT	2263	2264	HR -> QQ (IN REF. 1).
FT	CONFLICT	2289	2289	L -> V (IN REF. 1).
FT	CONFLICT	2347	2347	C -> S (IN REF. 1).
FT	CONFLICT	2426	2426	S -> R (IN REF. 1).
FT	CONFLICT	2609	2610	AV -> RC (IN REF. 1).
FT	CONFLICT	2684	2686	MISSING (IN REF. 1).
FT	CONFLICT	2756	2756	S -> T (IN REF. 1).
FT	CONFLICT	2897	2898	TA -> SP (IN REF. 1).
FT	CONFLICT	3026	3026	N -> P (IN REF. 1).
FT	CONFLICT	3097	3097	N -> F (IN REF. 1).
FT	CONFLICT	3272	3272	S -> A (IN REF. 1).
FT	CONFLICT	3317	3317	S -> R (IN REF. 1).
FT	CONFLICT	3452	3452	S -> Y (IN REF. 1).
FT	CONFLICT	3484	3486	HQMSHPFFETISH -> DEMSDAGLFRSE (IN REF. 1).
FT	CONFLICT	3500	3501	PH -> GQ (IN REF. 1).
FT	CONFLICT	3508	3508	E -> R (IN REF. 1).
FT	CONFLICT	3588	3588	AA; 402072 MW; BC02FB157D7F1FDB CRC64;
SO	SEQUENCE	3588	AA; 402072 MW; BC02FB157D7F1FDB CRC64;	

Query Match 4.7%; Score 95; DB 1; Length 3588;
Best Local Similarity 19.7%; Pred. No. 24;
Matches 63; Conservative 48; Mismatches 111; Indels 98; Gaps 13;

QY	14	VSQMAILAFFMYSHINISLSMKQAQPMHVLVLSWSRSGSFVQGLFGQHPDVFYLM	73
DB	990	VEKAGIFDNFFETGGHSLKXATLLTKHK	-----ETGIEIPQQLFEHTITALAE 1040
QY	74	PAWHVMTKQSTAWMLHMAVLDLIRAVFLCDMSVFDAYMEGPRGROSSLFWENSRALC	133
DB	1041	EADH	-----RESKAF-----AVIEPAKQEHYPLHWHSEHIS 1073
QY	134	SAPA	-----CDIIPQDEII-----PRAHCLLCS 157
DB	1074	SASSRMRESAISASSNSSEGFYSKAGARISGINTPRVIESFVLENSTPRQKIH	VCV 1132
QY	158	QQPFVVEKACRSYSHVVKVFRFNLSLYLLXDP	-----NLHIV-----HLVRDRP 208
DB	1133	DFNIEMIEGRSDEAIMASFVRTDLAKA	PLFRIGLGLLEENRMLLFDVHHLISDOV 1191
QY	209	AFPRSEERTKGLMID	-----SRIVMGQHEQKLKEEDOPYYVMQVICOQLFIYKTIQSLPKA 266
DB	1192	SI	-----GIMLEELARIYKGEQLDLRLQYKDYAVWQ-----SRQAAGYKKDQAYWKE 1240
QY	267	L-QERYLLVRVEDLARAPV	284
DB	1241	VFAGELEVLQLLSDYPRPV	1260

RESULT 5
TPSB_CAEEL STANDARD; PRT; 359 AA.
ID_TPSB_CAEEL

```
AC Q20351;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative protein-tyrosine sulfotransferase (EC 2.8.2.20)
DE (Tyrosylprotein sulfotransferase) (TPST).
GN F42G9.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Taich A.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the O-sulfation of tyrosine residues within
CC acidic motifs of polypeptides (By similarity).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + protein tyrosine =
CC adenosine 3',5'-bisphosphate + protein tyrosine-O-sulfate.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By
CC similarity).
CC -!- SIMILARITY: Belongs to the protein sulfotransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J00051; AAA91354.1;
CC F42G9.8; T16350; T16350.
CC WormPep; F42G9.8; CE07235.
CC InterPro; IPRO00863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC KW Hypothetical protein; Transferase; Transmembrane; Glycoprotein;
CC Signal-anchor.
FT DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 28 48 (SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 49 359 LUMENAL, CATALYTIC (POTENTIAL).
FT SEQUENCE 359 AA; 41469 MW; 9FC2F44539B8CC53 CRC64;
SQ
Query Match 4.5%; Score 92; DB 1; Length 359;
Best Local Similarity 19.5%; Pred No. 2.4;
Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;
QY 52 RGSFVGLFQGHDPV----PYLMEPAHVMVMTKQSTAWMLHMAVRLIRAVFLCDMS 107
DB 95 RSGTTLMAILDAHPDRCGGETMLLPSTWQAGWRND-WYNNSGI-----TQE 143
QY 108 VPDAYMEPGFRCSLFWQNSRALCSAPACDIIPQ-DEIIPRAHCRLLCSQQPFVEVK 166
DB 144 VFDDAV-----SAITIVAGHSELAPR-----LCNKDP----- 172
QY 167 ACSYSHVYLKVEYRFFNLQSLYPLKLPDSLNLIHVLVRDPRAVRSRRTK-----GDLN 222
DB 173 ----YFALMPTIR-----RLYP-----NAKFLMIRDAZAVVHSMIERKVPVAGYNT 216
QY 223 IDSRIWVGHEQKLKEDQPYVMQVICSQ-----LEIVKTIQSLPKALQERYLLVRVED 278
DB 217 SDEISMFPVQNNQELRK-----MTFCNNAPQCCKIVY-----YER 251
QY 279 LAPAPVAQTSRMVYFVGLFLEPLPLHPLQTVWVNIIRGKMGDGHAFH-TNARDALNVQAWRWS 337
DB 252 LIQKPAEILRIITNFDLPFSQQMLRHQDLIGVDLNDQFSAQVKSINTKALTSWF 311
QY 338 LPY-EKVSRLQKAGDAMLLGY 359
DB 312 DCFSEETLRKLDVAPFLGILGY 334
```

```
RESULT 6
MACF HUMAN STANDARD; PRT; 5430 AA.
AC Q9UPN3; Q75053; Q8WXY2; Q9H540; Q9UKP0; Q9ULG9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Microtubule-actin crosslinking factor 1, isoforms 1/2/3 (Actin cross-
DE linking family protein 7) (Macrophin 1) (Trabeculin-alpha) (620 kDa
DE actin-binding protein) (ABP620).
GN MACF1 OR ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
EN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=20001959; PubMed=10529403;
RA Okuda T., Matsuda S., Nakatsugawa S., Ichigotani Y., Iwahashi N.,
RA Takahashi M., Ishigaki T., Hamaguchi M.,
RT "Molecular cloning of macrophin, a human homologue of Drosophila
RT kakapo with a close structural similarity to plectin and dystrophin.";
RL Biochem. Biophys. Res. Commun. 264:568-574(1999).
EN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20026884; PubMed=10559237;
RA Sun Y., Zhang J., Kraeft S.-K., Auclair D., Chang M.-S., Liu Y.,
RA Sutherland R., Saigia R., Griffin J.D., Ferland L.H., Chen L.B.;
RT "Molecular cloning and characterization of human trabeculin-alpha, a
RT giant protein defining a new family of actin-binding proteins.";
RL J. Biol. Chem. 274:33522-33530(1999).
EN [3]
RP SEQUENCE OF 182-4812 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=21833812; PubMed=11845288;
RA Gong T.-W.L., Bestirli C.G., Lonax M.I.;
RT "XACFI gene structure: a hybrid of plectin and dystrophin.";
RL Mamm. Genome 12:852-861(2001).
EN [4]
RP SEQUENCE OF 868-2350 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirosewa M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
EN [5]
RP SEQUENCE OF 1544-5057 FROM N.A.
RA Corby N.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
EN [6]
RP SEQUENCE OF 3312-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
EN [7]
RP SEQUENCE OF 3734-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira N., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakajima D., Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
CC -!- FUNCTION: F-actin-binding protein which may play a role in cross-
CC linking actin to other cytoskeletal proteins. Also binds to
CC microtubules (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
```


QY 174 -VVLKEVFFNLSQYPLIKPSLNLHIVLRDPRVFRSRTKGLMIDSR----- 226
 Db 1874 FWTISQKVLDMENFKEGKEPS-----ETGNLVKD-----KUKDATERVTAHLSKCTRLGS 1925
 QY 227 ---IYMQHEQKXKEDQFYVMQVICSQLE--IYKTIQSLPKALQERYLLVRYEDLAR 281
 Db 1926 HLMMLGQVHQFQNSGADSLQAMQA-CEANVEKLLSDTVASDPGVLSQ----- 1973
 QY 282 APVAOTSMYEVGVLEFLP--HLOTWVNI--TRKGMGDHAFHTNAEDA-LNVYQAWRW 336
 Db 1974 --LATTKQLELAHQVPEKLVKQKARDIMEIEGEPADPHRHVQVETDTSILSHFQSLSY 2031
 QY 337 SLPEYKVSRLQKACDAMNLLGYRVSREQEQRNLLL 373
 Db 2032 SLA-ERSLILQXIAQS-----QSVQESLESLL 2058

RESULT 7
 MAC4_HUMAN
 ID MAC4_HUMAN STANDARD; PRT; 5938 AA.
 AC Q96PK2; Q8WXY1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Microtubule-actin crosslinking factor 1, isoform 4.
 GN MACF1 OR ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Buthera; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RX MEDLINE=21833812; PubMed=11845288;
 RA Gong T.-W.L., Besirli C.G., Lomax M.I.;
 RT "MACF1 gene structure: a hybrid of plectin and dystrophin.";
 RL Mamm. Genome 12:852-861(2001).
 CC -!- FUNCTION: May play a role in cross-linking cytoskeletal proteins
 CC by binding intermediate filaments to the N-terminal plectin
 CC repeats and microtubules to the C-terminus.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=4;
 CC IsoId=Q96PK2-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q9UPN3-1; Sequence=External;
 CC Name=2;
 CC IsoId=Q9UPN3-2; Sequence=External;
 CC Name=3;
 CC IsoId=Q9UPN3-3; Sequence=External;
 CC -!- TISSUE SPECIFICITY: Expressed in heart, lung, pituitary and
 CC placenta, not found in brain, kidney, liver, pancreas or skeletal
 CC muscle.
 CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -!- SIMILARITY: Contains 19 plectrin repeats.
 CC -!- SIMILARITY: Contains 32 spectrin repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF317696; AAL09459.1; -;
 DR EMBL; AF325341; AAL38997.1; -;
 DR EMBL; AF325333; AAL38997.1; JOINED.
 DR EMBL; AF325334; AAL38997.1; JOINED.
 DR EMBL; AF325335; AAL38997.1; JOINED.
 DR EMBL; AF325336; AAL38997.1; JOINED.

DR EMBL; AF325339; AAL38997.1; JOINED.
 DR EMBL; AF325340; AAL38997.1; JOINED.
 DR GO; GO:0005856; Cytoskeleton; ISS.
 DR GO; GO:0005509; F:calcium ion binding; ISS.
 DR GO; GO:0008017; F:microtubule binding; ISS.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR003108; GAS2.
 DR InterPro; IPR001101; Plectrin repeat.
 DR InterPro; IPR02017; Spectrin.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF02187; GAS2; 1.
 DR Pfam; PF00681; Plectrin; 11.
 DR Pfam; PF00435; spectrin; 26.
 DR ProDom; PD000012; EF-hand; 1.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00243; GAS2; 1.
 DR SMART; SM00250; PLEC; 19.
 DR SMART; SM00150; SPEC; 32.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR Cytokeleton; Calcium; Calcium-binding; Repeat; Coiled coil;
 KW Alternative splicing.
 FT DOMAIN 1830 1936 COILED COIL (POTENTIAL).
 FT DOMAIN 2001 2192 COILED COIL (POTENTIAL).
 FT DOMAIN 2282 2345 COILED COIL (POTENTIAL).
 FT DOMAIN 2477 2507 COILED COIL (POTENTIAL).
 FT DOMAIN 2541 2654 COILED COIL (POTENTIAL).
 FT DOMAIN 2686 2814 COILED COIL (POTENTIAL).
 FT DOMAIN 2887 2919 COILED COIL (POTENTIAL).
 FT DOMAIN 3046 3197 COILED COIL (POTENTIAL).
 FT DOMAIN 3262 3503 COILED COIL (POTENTIAL).
 FT DOMAIN 3632 3666 COILED COIL (POTENTIAL).
 FT DOMAIN 3746 3779 COILED COIL (POTENTIAL).
 FT DOMAIN 3920 3984 COILED COIL (POTENTIAL).
 FT DOMAIN 4098 4168 COILED COIL (POTENTIAL).
 FT DOMAIN 4288 4308 COILED COIL (POTENTIAL).
 FT DOMAIN 4354 4386 COILED COIL (POTENTIAL).
 FT DOMAIN 4397 4433 COILED COIL (POTENTIAL).
 FT DOMAIN 4469 4489 COILED COIL (POTENTIAL).
 FT DOMAIN 4586 4720 COILED COIL (POTENTIAL).
 FT DOMAIN 4845 4880 COILED COIL (POTENTIAL).
 FT DOMAIN 4910 4939 COILED COIL (POTENTIAL).
 FT DOMAIN 4970 5000 COILED COIL (POTENTIAL).
 FT DOMAIN 5409 5437 COILED COIL (POTENTIAL).
 FT DOMAIN 5546 5569 COILED COIL (POTENTIAL).
 FT CA_BIND 5598 5610 EF-HAND 1 (POTENTIAL).
 FT CA_BIND 5634 5646 EF-HAND 2 (POTENTIAL).
 FT REPEAT 12 49 PLECTIN 1.
 FT REPEAT 53 88 PLECTIN 2.
 FT REPEAT 89 126 PLECTIN 3.
 FT REPEAT 130 164 PLECTIN 4.
 FT REPEAT 166 202 PLECTIN 5.
 FT REPEAT 203 240 PLECTIN 6.
 FT REPEAT 243 278 PLECTIN 7.
 FT REPEAT 279 316 PLECTIN 8.
 FT REPEAT 318 354 PLECTIN 9.
 FT REPEAT 725 762 PLECTIN 10.
 FT REPEAT 763 800 PLECTIN 11.
 FT REPEAT 801 838 PLECTIN 12.
 FT REPEAT 839 876 PLECTIN 13.
 FT REPEAT 897 934 PLECTIN 14.
 FT REPEAT 935 972 PLECTIN 15.
 FT REPEAT 975 1011 PLECTIN 16.
 FT REPEAT 1012 1049 PLECTIN 17.
 FT REPEAT 1121 1158 PLECTIN 18.
 FT REPEAT 1159 1196 PLECTIN 19.
 FT REPEAT 1925 2032 SPECTRIN 1.
 FT REPEAT 2052 2160 SPECTRIN 2.
 FT REPEAT 2211 2313 SPECTRIN 3.
 FT REPEAT 2320 2430 SPECTRIN 4.
 FT REPEAT 2437 2543 SPECTRIN 5.
 FT REPEAT 2550 2652 SPECTRIN 6.
 FT REPEAT 2659 2758 SPECTRIN 7.
 FT REPEAT 2765 2896 SPECTRIN 8.

```

FT REPEAT 2903 3008 SPECTRIN 9.
FT REPEAT 3015 3119 SPECTRIN 10.
FT REPEAT 3126 3229 SPECTRIN 11.
FT REPEAT 3236 3339 SPECTRIN 12.
FT REPEAT 3346 3446 SPECTRIN 13.
FT REPEAT 3453 3555 SPECTRIN 14.
FT REPEAT 3562 3664 SPECTRIN 15.
FT REPEAT 3671 3775 SPECTRIN 16.
FT REPEAT 3782 3884 SPECTRIN 17.
FT REPEAT 3891 3993 SPECTRIN 18.
FT REPEAT 4000 4102 SPECTRIN 19.
FT REPEAT 4109 4211 SPECTRIN 20.
FT REPEAT 4218 4320 SPECTRIN 21.
FT REPEAT 4327 4428 SPECTRIN 22.
FT REPEAT 4438 4544 SPECTRIN 23.
FT REPEAT 4551 4653 SPECTRIN 24.
FT REPEAT 4660 4763 SPECTRIN 25.
FT REPEAT 4770 4872 SPECTRIN 26.
FT REPEAT 4879 4982 SPECTRIN 27.
FT REPEAT 4989 5091 SPECTRIN 28.
FT REPEAT 5098 5201 SPECTRIN 29.
FT REPEAT 5208 5309 SPECTRIN 30.
FT REPEAT 5316 5418 SPECTRIN 31.
FT REPEAT 5425 5555 SPECTRIN 32.
FT CONFLICT 1712 1712 P -> S (IN REF. 1; AAL38997).
SQ SEQUENCE 5938 AA; 670134 MW; B8784112752DA004 CRC64;

Query Match
Best Local Similarity 22.2%; Pred. No. 87; Length 5938;
Matches 88; Conservative 66; Mismatches 128; Indels 115; Gaps 24;

QY 2 2 LIPKMKLLFLVVSQMAILALFFHMYSHNISLSMKAQPERHVLVSSWRSQSFVQQL 61
Db 2254 LLSQQQNFIATQSAQA-----FLDQGHNL-----PEEQML-----QQKLGL 2294

QY 62 FQCHDEVFLMEPAHVMTFQSTANLHM-AVDLIRAVFLCDMSVDFVMEPGPRQ 120
Db 2295 KEQYS-----TSLAQSAELKQVTLQDELQK-FLQHKKEFSWLE---RSE 2337

QY 121 SLSFOWENSRLCSAPACDIIPQDEIIPRAHRLCSQPPF--EYVEKACRSYSH- 173
Db 2338 KEL---ENMHKGGSP-----ETLP-----SLLRKQGSFSDVI-----SHKGLR 2375

QY 174 -VVLKVEVFNLSLYPLKDSMLNHLVLRDPAVRSRERKGLMIDSR----- 226
Db 2376 FVTISQKVLWNSFKGKES-----EIGNLYKD-----KLKDATERVTLHSKCTRUGS 2427

QY 227 ---IVNGQHEQKLKEDQPYVWQVVCOSQLF--IYKTIQSLPKALQERYLLVRYEDLAR 281
Db 2428 HLNMLLGQYHQFQNSADSLOAMQA-CEANVEKLSDIVASDPGVLOEQ----- 2475

QY 282 AVPAQTSRMYEFGVLEFLP--HLQTVHNI--TRKGMGDGHAFHTNARDA-LNVSQAWRW 336
Db 2476 --LATTQQLQELAEHQVVEKLVQKQVARDIMEIEGAPDHRHQVQETTDLSLHPQSLSY 2533

QY 337 SLPEYKVSRLQACGADNALLGLVHRSRQEQRNLL 373
Db 2534 SLA-ERSLLQKRAIQS-----QSQESLESLL 2560

RESULT 8
ID 239F HUMAN STANDARD; PRT; 294 AA.
AC Q1577;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fetal brain protein 239 (239Fb).
GN C11ORF8 OR FAM1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;

```

```

RN [1] SEQUENCE FROM N.A.
RX MEDLINE=95080775; PubMed=7527372;
RA Schwartz F., Neve R., Eisenman R., Gessler M., Bruns G.;
RT "A WAGR region gene between PAX-6 and FSHB expressed in fetal brain.";
RL Hum. Genet. 94:658-664(1994).
RN [2] SEQUENCE FROM N.A.
RX MEDLINE=96115606; PubMed=8666403;
RA Schwartz F., Eisenman R., Knoll J., Gessler M., Bruns G.;
RT "cDNA sequence, genomic organization, and evolutionary conservation
of a novel gene from the WAGR region.";
RL Genomics 29:526-532(1995).
RN [3] SEQUENCE FROM N.A.
RC TISSUE-Colon, Kidney, and Stomach;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshyuk S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., J. J., Hulyk S.W.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN FETAL BRAIN.
CC -!- SIMILARITY: BELONGS TO THE UPF0046 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; U57911; AAC50564.1; --
DR EMBL; BC031582; AAH31582.1; --
DR Genbank; HGNC:1180; C11orf8.
DR MIM; 600911; --
DR GO; GO:0007399; P:neurogenesis; TAS.
DR InterPro; IPR004843; M-ppestrase.
DR Pfam; PF00149; Metallophos; 1.
SQ SEQUENCE 294 AA; 33360 MW; 43B2BC0DA1BFD1F0 CRC64;

```

```

Query Match
Best Local Similarity 20.2%; Pred. No. 2; Length 294;
Matches 53; Conservative 30; Mismatches 89; Indels 90; Gaps 9;

QY 139 DIIPQDEIIPRAHRLCSQPPFVEVKEACRSYSHVL-----KEVRFNQLSL 187
Db 45 DIPYDTPKPAHTRFVCISDTHSRDTGQMPYGDILLHTGDTFELGLPSEVKFN----- 100

QY 188 YPLLKQPSLNLHVLVRDPAVRSRERKGLMIDSRVM-GQHEQKLKKE-----D 240
Db 101 -----DWLGNLPYEYKIVLGNHLELFDKEFMADLVK 132

QY 241 QPYVWQVICOQSLIYKTIQSLPKALQERYLLVRYEDLARPAVQTSRM-----YEFVGL 296
Db 133 QDYVRRPVSVKLPEDPDNVQSL-----LTNSIYLODSEVTVKGFRIYGA 177

```

```

QY 297 EFLPHLQTVHNIITGCKMGDHAFTNARDALNVSQAWSLPYEKYSRLOKACGDAMNL 356
DB 178 PWTWFGNGFNLPKQGLLD-----KWNLIPEGIDILM-THGPP--- 216
QY 357 LGYHRVSEGEQRNLLDLST 378
DB 217 LGFRDWVPELQVGCVELLNT 238

RESULT 9
MX3_RAT
ID MX3_RAT STANDARD; PRT; 659 AA.
AC P18590;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Interferon-induced GTP-binding protein MX3.
GN MX3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=91056598; PubMed=2173790;
RA Meier E., Kunz G., Haller O., Arnheiter H.;
RT "Activity of rat Mx proteins against a rhabdovirus.";
RL J. Virol. 64:6263-6269(1990).
CC -!- FUNCTION: Does not show activity against influenza virus or VSV;
CC although it only differs from Mx2 by 8 positions.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- INDUCTION: By interferons.
CC -!- SIMILARITY: Belongs to the dynamin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52713; CAA36937.1; -
DR PIR; S11737; S11737.
DR InterPro; IPR001401; Dynamin.
DR InterPro; IPR000375; Dynamin_central.
DR InterPro; IPR003130; GED.
DR Pfam; PF00350; dynamin; 1.
DR Pfam; PF01031; dynamin_2; 1.
DR Pfam; PF02212; GED; 1.
DR PRINTS; PR00195; DYNAMIN.
DR SMART; SM00053; DINC; 1.
DR SMART; SM00302; GED; 1.
DR PROSITE; PS00410; DYNAMIN; 1.
KW Interferon induction; GTP-binding; Multigene family.
FT NP_BIND 75 82 GTP (POTENTIAL).
FT NP_BIND 176 180 GTP (POTENTIAL).
FT NP_BIND 245 248 GTP (POTENTIAL).
SQ SEQUENCE 659 AA; 74951 MW; F55D63283BC865B3 CRC64;

Query Match 4.5%; Score 91; DB 1; Length 659;
Best Local Similarity 20.6%; Pred. No. 6.2;
Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;

QY 62 FGQHPDVFMPEAWVWMTFKQSTAWMLHMAVRDLI-----RAVFLCDMSVFDAWNE--- 114
DB 301 FKEHPQFALLEDG-----RATVPECLAERTMTLISHICKSLPLENQIKESHQSTSE 353
QY 115 -----PGPRQSSLFOWNSRALCSAPACDIIPODEIIPRAHCRLLCS-QQPF--- 161
DB 354 ELQKYGADIPEDENEXTLFLEKINAF-NQDITAVEGEIYREKCRFLFTKLKKEFFLW 412
QY 162 -EVVEKACRSYSHVWVKEVRFNLIQ-----SLYPLLKDPSLNLIH 200

```

```

DB 413 SEIERNFQKGSALYKEVYIFEMQYRGRELPGFVNYKTFENIIRQIKTLEEPAMEM-- 470
QY 201 VHLVRD-PAVPR-----SRRTKGLDMLTDSIRVMQHQKLEKEDQPYVMQV 248
DB 471 LHKVTEIVRAAFTTVSEKNFSEFFNLHRTTKLEDIRL---EQETEAKAIRLHFQMEQ 527
QY 249 ICQSLEIYKTIQSLPKAQERYLLVRYEDL-----ARAPVAQTSRMVYFVGL 296
DB 528 IYQCQDIYR-----KALQK-----VREEEAEEERKHKGSRSQAQSPNLQTSMD----- 572
QY 297 EFLPHLQTVHNIITGCKMGDHAFTNARDALNVSQAWSLPYEKYSRLOKACGDAMNL 356
DB 573 EIPQHLNAYRQBAHN 587

RESULT 10
C7D8_SOYBN
ID C7D8_SOYBN STANDARD; PRT; 504 AA.
AC O81974;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 71D8 (EC 1.14.-.-) (P450 CP7).
GN CYP71D8.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eusoids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=cv. Harosoy 63;
RX MEDLINE=98311068; PubMed=9648734;
RA Schopfer C.R., Ebel J.;
RT "Identification of elicitor-induced cytochrome P450s of soybean
RT (Glycine max L.) using differential display of mRNA.";
RL Mol. Gen. Genet. 258:315-322(1998).
CC -!- INDUCTION: By fungal elicitor.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y10493; CAA71517.1; -
DR PIR; T07120; T07120.
DR HSP; P14779; LJPZ.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Heme.
FT METAL 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 504 AA; 57579 MW; 86E4FAEBC2FF2A8F CRC64;

Query Match 4.4%; Score 89.5; DB 1; Length 504;
Best Local Similarity 19.9%; Pred. No. 5.9;
Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16;

QY 68 VFYLMPEAWVWMTFKQSTAWMLHMAVRDLIIRAVFLCDMSVFDAWMEPGPRQSSLFOWE 127
DB 14 VFLLH--WLIV-KTYKQKSHK-----PPGWR----- 39
QY 128 NSRALCSAPACDIIPODEIIPRAHCRLLCSQQPFVVEKACRSYS---HVVLKEVRFNLI 184
DB 40 -----LP-----IIGNLHQLAALASLPDQALQKLVKRYGPLMLQLQGEISTLV 83
QY 185 QSLYPLLKDPSLNLIHVIHLVRDPAVPRSRERIKGDMIDSRVMQHQKLEKEDQPY 244

```

```

Db 84 SS--PKMAMENKTHDVHFORPO-----LLAQFMYVYGTDTAFAPYGDYWR 129
QY 245 VMQVVCOSQLEIYKTIQSLPALQE--RYLLVRVEDLARAPVACTSMRYEFGVL----- 296
Db 130 QIRKICTLELSAKRVOSFHSIRODENKKTQSIHSSAGSPIDLSGKFLSLGTTVSRAA 189
QY 297 -----EFL-----PHLOTWVHNITRKGMDGDAFTNA----- 324
Db 190 FGKENDDODEFMSLVRAKAITWGCGFEVDDMEPSLKP-LHLLTROKAKVHV-HQRADKIL 247
QY 325 RDALNVSOAWRWSLPYKRVSLQKACGDAMNLLGYRHRVRSFQBNRLLDL 376
Db 248 EDILRKHM-----EKRTVKSGNG-----SEAEQEDLVLL 279

```

RESULT 11

```

PKSM_BACSU
ID PKSM_BACSU STANDARD; PRT; 4273 AA.
AC P40872; O31781;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative polyketide synthase pksm.
CN PKSM OR PKSY OR BSU7200.
OS Bacillus subtilis.
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=9804033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Boilotin A., Borchert S.,
RA Borriess R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Fittington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klarr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Taseuchi M., Tamakoshi A., Tanaka T., Terptrat P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
RN [2]
RP SEQUENCE OF 1-1763 FROM N.A.
RC STRAIN=168 / PB1424;
RA Tognoni A., Grandi G.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -!- COFACTOR: Contains 4 covalently bound phosphopantetheines
CC (Potential).

```

```

CC -!- SIMILARITY: Contains 4 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z99113; CAB13603.1; -
CC EMBL; Z35133; CAB44505.1; -
CC PIR; C69679; C69679.
CC Subtilisin, BG10931; pksm.
CC InterPro; IPR002198; ADH short.
CC InterPro; IPR000794; Ketoacyl synth.
CC InterPro; IPR001601; Methyltransf.
CC InterPro; IPR006163; Pp bind.
CC InterPro; IPR006162; Pantine S.
CC InterPro; IPR000051; SAM bind.
CC Pfam; PF00106; adh_short; 1.
CC Pfam; PF00109; ketoacyl-synt; 3.
CC Pfam; PF02801; ketoacyl-synt_C; 3.
CC Pfam; PF00550; pp-binding; 4.
CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
CC PROSITE; PS00806; B-KETOACYL SYNTHASE; 2.
CC PROSITE; PS00875; ACP_DOMAIN; 4.
KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 295 364
FT ACYL CARRIER (ACP) 1.
FT BETA-KETOACYL SYNTHASE 1.
FT DOMAIN 396 834
FT ACYL CARRIER (ACP) 2.
FT BETA-KETOACYL SYNTHASE 2.
FT DOMAIN 2190 2258
FT BETA-KETOACYL SYNTHASE 3.
FT DOMAIN 3532 3947
FT ACYL CARRIER (ACP) 3.
FT DOMAIN 3410 3483
FT ACYL CARRIER (ACP) 4.
FT DOMAIN 4140 4209
FT PHOSPHOPANTHETHEINE (POTENTIAL).
FT BINDING 327 327
FT BINDING 2222 2222
FT ACT SITE 2476 2476
FT BINDING 3446 3446
FT ACT SITE 3690 3690
FT BINDING 4172 4172
FT CONFLICT 103 103
FT CONFLICT 276 276
FT CONFLICT 289 289
FT CONFLICT 289 289
SQ SEQUENCE 4273 AA; 477459 MW; 3BBFCF1A250AEB5A CRC64;
Query Match 4.4%; Score 89.5; DB 1; Length 4273;
Best Local Similarity 22.9%; Pred.No. 85;
Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;
QY 161 FEVVEKACRSYSHVWLKEVFFENLQSLYPLKDPSS---LNLHIVHLVRDPAVRSRER 216
Db 36 QYVFQEHGAYQOELELK-----NLTFYPLADESYDIALTIHVSEEGTWSIIIDGQK 90
QY 217 TKGDLMDISRIYVNGQHE-QKLKKEQPYVMQVICOQLEIKYKTIQSLPALQERYLLVR 275
Db 91 QHGESLSDKR---QYETADMRKQETAFAESI-----DLNOWNKSTADRLNLDIYEOCR 142
QY 276 YEDLARAPVAQT-SRMAYE-----FVGLFELPHLOTWVHNIT--RGKGMGDHAFHTN 323
Db 143 SQELVHTGMKREGQIYEAKGAVIDLAVGQALRHSDAFLPHPTLIDSGIGS----- 196
QY 324 ARDALNVSOAWRWSLPYKRVSLQKAC-----RLQKAC 350
Db 197 --SCLISDQTMVLPVLYESFSASERLQKGC 224
RESULT 12
P4ST FLACH
ID P4ST FLACH STANDARD; PRT; 320 AA.
AC P52837;
DT 01-OCT-1996 (Rel. 34, Created)

```


DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Flavonol 4'-sulfotransferase (EC 2.8.2.-) (F4-ST).
 OS Flaveria chloraefolia.
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Asterales; Asteraceae; Asteroideae; Tageteae; Flaveria.
 OX NCBI_TaxID=4228;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92159034; PubMed=1741382;
 RA Varin L., Deluca V., Ibrahim R.K., Brisson N.;
 RT "Molecular characterization of two plant flavonol sulfotransferases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).
 RN [2]
 RP PAPS-BINDING SITE.
 RX MEDLINE=95279378; PubMed=7759495;
 RA Varin L., Marsolais F., Brisson N.;
 RT "Chimeric flavonol sulfotransferases define a domain responsible for
 RT substrate and position specificities.";
 RL J. Biol. Chem. 270:12498-12502(1995).
 CC -!- FUNCTION: Transfers sulfate group into flavonol position 4'.
 CC May play a role in auxin transport.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: Highest in shoot tips and lowest in mature
 CC leaves and roots (By similarity).
 CC -!- SIMILARITY: Belongs to the plant sulfotransferase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M841136; AAA33343.1;
 DR HSSP; P50224; 1CJM.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 KW Transferase.
 FT BINDING 148 163 PAPS-BINDING SITE.
 SQ SEQUENCE 320 AA; 37255 MW; 5B263659F6CCBCC0 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 320;
 Best Local Similarity 20.8%; Pred. No. 3.7;
 Matches 56; Conservative 30; Mismatches 83; Indels 100; Gaps 12;
 QY 54 GSSEVQLFGQHP-DVEYLMERPAWHVMTFKOSTAWMLHMAVRDLIRAVFLCDMSVEDAY 112
 DB 47 GALLAQSPKARDDDVFLCSYP-----KSGTWLKLALAVIVR-----EKDFEF 91
 QY 113 MEP-----GPRQSSLFQ-WENSRALCAPACDIIIPQDEIIPRAHCRLLC-----SQPFVEVE 165
 DB 92 TSPLLATNPINCIPYIEKDKKIVENQNSCTP-----MATHMPYH 133
 QY 163 VVEKACRSYSHVVLKEVRFPNLSQLYPLKDPNLNHLVHVRDPAVRSRERTKGDLM 222
 DB 134 VLPS-----IL-----ALNCKMVIYTRNKDV----- 156
 QY 223 IDSRIYMGQHEQKLKEDQPYVMQVVICOSQLEIYKTIOSLP-----KALQER-- 270
 DB 157 IVSFYHFGREITKLPEDAPFE-----EAFDEFYHGISQFGPYWDHLLGYKASLERPE 210
 QY 271 -YLLVYEDLARAPVACTSMYFVCLFEP 298
 DB 211 VILFLKYEDVKDPTSNVRLAEFIGYFP 239
 RESULT 13
 ID MPPA_RAT STANDARD; PRT; 524 AA.

P20069;
 01-FEB-1991 (Rel. 17, Created)
 01-FEB-1991 (Rel. 17, Last sequence update)
 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mitochondrial processing peptidase alpha subunit, mitochondrial
 DE precursor (EC 3.4.24.64) (Alpha-MPP) (P-55).
 GN MPPCA OR INP5E OR MPPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=91045920; PubMed=2236012;
 RA Kleiber J., Kalousek F., Swarcop M., Rosenberg L.E.;
 RT "The general mitochondrial matrix processing protease from rat liver:
 RT structural characterization of the catalytic subunit.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7978-7982(1990).
 CC -!- FUNCTION: The mitochondrial processing protease (MPP-I) cleaves
 CC presequences from mitochondrial protein precursors. Most MPP-I
 CC cleavage sites follow an arginine at position -2.
 CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
 CC precursor proteins imported into the mitochondrion, typically with
 CC Arg in position P2.
 CC -!- SUBUNIT: Heterodimer of alpha and beta subunits.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- SIMILARITY: Belongs to peptidase family M16.
 CC -!- CAUTION: Does not seem to have a protease activity as it lacks the
 CC zinc-binding site.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M57728; AAA41632.1;
 DR PIR; A36205; A36205.
 DR MEROPS; M16.971;
 DR InterPro; IPR001431; Peptidase M16.
 DR InterPro; IPR007863; Peptidase_M16_C.
 DR Pfam; PF00675; Peptidase_M16; 1.
 DR Pfam; PF05193; Peptidase_M16_C; 1.
 DR PROSITE; PS00143; INSULINASE; 1.
 KW Hydrolyase; Metalloprotease; Mitochondrion; Transit peptide.
 FT TRANSIT 1 32 MITOCHONDRION.
 FT CHAIN 33 524 MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA
 SUBUNIT
 SQ SEQUENCE 524 AA; 58607 MW; 8BF08FBC9FF09DB2 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 524;
 Best Local Similarity 21.5%; Pred. No. 6.8;
 Matches 70; Conservative 46; Mismatches 104; Indels 106; Gaps 19;
 QY 70 YLM--EPAW---HVMTFKQ-----STAWMLHMAVRDLIRAVFLCDMSVEDAYMEPG 116
 DB 269 YLGVQPAWGAQAVNWLTAQWHTGGSRW-----RE-----TCMSA-----LRP- 311
 QY 117 PRQSSLFQWENSRALCAPACDIIIPQDEIIPRAHCRLLC-----SQPFVEVE 165
 DB 312 PRQSSHI-YGGAREL-----LLEEDFIPPAVLNMMWGGGSGSAGPGKGMFS--- 360
 QY 166 KACRSYSHVVLKEVRFPNLSQLYPLKDPNLNHLVHVRDPAVRSRERTKGDLMIDS 225
 DB 361 ---RLYNLVNRRHHWMYNATSYHSYED--TGKLCIHASADPROV-----REMVETITKE 410
 QY 226 RIVMGQHEQKLKEDQPYVMQVVICOSQLEIYKTIOSLPKALQERYLLVRYEDLARAPVA 285
 DB 411 FILMGRTVDLVEERAKTQLMSLMN-----LESRP-----VIFEDVGQVLA 454

```

QY 286 QTSRMVYFVGLFELPH-LQTVHVNITRGKMGDHPHTNARDALNVQAWWSLPYKVS 344
Db 455 TSHRK-----LPHELCTLRNV-----KPEDIKRVASKMLRGKP----- 488
QY 345 RLQKACGDAMNLLGYRHVRSQEQRN 370
Db 489 -AVALGDLTDLPTVEHQAALSSRD 513

RESULT 14
CAP STRPY STANDARD; PRT; 920 AA.
AC Q8A0U7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC).
GN PPC OR SPY0608.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -!- FUNCTION: Through the carboxylation of phosphoenolpyruvate (PEP)
CC it forms oxaloacetate, a four-carbon dicarboxylic acid source for
CC the tricarboxylic acid cycle.
CC -!- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +
CC phosphoenolpyruvate + CO(2).
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SIMILARITY: Belongs to the PEPCase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006516; AK33584.1; -
CC HSSP; P00884; LFY.
CC HAMAP; MF00595; -; 1.
CC InterPro; IPR001449; PEPCase.
CC Pfam; PF00311; PEPCase; 1.
CC PRINTS; PR00150; PEPCARBLASE.
CC PROSITE; PS00781; PEPCASE_1; 1.
CC PROSITE; PS00393; PEPCASE_2; 1.
CC Lyase; Carbon dioxide fixation; Tricarboxylic acid cycle;
CC Complete proteome.
CC ACT_SITE 138 138 BY SIMILARITY.
FT ACT_SITE 583 583 BY SIMILARITY.
SQ SEQUENCE 920 AA; 104751 MW; 464EA4A309A22237 CRC64;

Query Match 4.4%; Score 89; DB 1; Length 920;
Best Local Similarity 18.6%; Pred. No. 14;
Matches 49; Conservative 48; Mismatches 85; Indels 82; Gaps 15;

QY 157 SQPF-----EVEKACRSYSHVVLKEVRFNQLSFLYLLKDPSLNLIHVL 203
Db 634 TSQPFSGIKDRILRTQGGIIE---NKYGN---KDVAYVHLE---MLISASINRMVQM 683
QY 204 VRDPAVFSRRTKGLMIDSIYVQGEQKKEKEDOPYVYVQVICOQLEIYKTIQSL 263
Db 684 ITDFNEIDSFRE-IMDSIVADSNIIY-----RKL-VFDNPHFYDFEASPI-----KEVSSL 734

```

```

QY 264 PRALQERYLLVRYEDLAPARVAQTSRMVYFVGLFELPHLQTVHVN----- 308
Db 735 -----NIGSRPAARKT-ITETGLRAIPWVFSWQNRIMFPGWYGVGSAPK 779
QY 309 --ITRGKMGD---HAFHTNARDALNVQAWWSLPYKVSRLQKACGDA-MNL-LGYRH 361
Db 780 RYIDRAQGNLERLQHYQT-----WPFHSLLSNVDMVLSKSNNTAFQYAQ 826
QY 362 VRSEQEQRNLLDLLSTWTVPQI 385
Db 827 LAERQDRVDVFEILDWOLTKNV 850

RESULT 15
RRPO_PVMR STANDARD; PRT; 1968 AA.
AC P17965; Q89548;
DT 01-NOV-1990 (Rel. 16, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA replication protein (147 kDa protein) (ORF 1) [Contains: RNA-
DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus M (strain Russian) (PVM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;
OC Carlavirus.
OX NCBI_TaxID=12168;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91116326; PubMed=1990070;
RA Zavrjev S.K., Kanyuka K.V., Levay K.E.;
RT "The genome organization of potato virus M RNA."
RL J. Gen. Virol. 72:9-14(1991).
CC -!- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SIMILARITY: Contains 1 OTU domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14449; BAA03339.1; -
CC PIR; P00093; PNO093.
CC MEROPS; C23.001; -
CC InterPro; IPR003323; OTU.
CC InterPro; IPR008041; Peptidase_C23.
CC InterPro; IPR001788; RNA_dep_RNapol2.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC InterPro; IPR006066; Viral_helicase1.
CC Pfam; PF02338; OTU; 1.
CC Pfam; PF05379; Peptidase_C23; 1.
CC Pfam; PF00978; RNA_dep_RNapol2; 1.
CC Pfam; PF01443; Viral_helicase1; 1.
CC PROSITE; PS50802; OTU; 1.
CC ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
CC Transferase.
CC DOMAIN 883 991 OTU.
FT NP_BIND 1166 1173 ATP (POTENTIAL).
SQ SEQUENCE 1968 AA; 223384 MW; 6F15A79E1AD96AAC CRC64;

Query Match 4.4%; Score 89; DB 1; Length 1968;
Best Local Similarity 20.7%; Pred. No. 36;
Matches 71; Conservative 38; Mismatches 108; Indels 126; Gaps 18;

QY 23 FFHYSHNTSSLKQAPRMHVLVLSWRSGSSFVQGLFGQHPDV----- 68

```

Db 367 FARLVIHN-SSWCATIMPEQL-----KEFMGNWLGKMPSVLARRFSSVRACVN 414
QY 69 -----FYLMEPAWHVMTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPG 116
Db 415 KFIRGLKEYSFTLRLEITW--WNIWENSYAWF-----FDTDAEVD 453
QY 117 -PRROSSLFOWENS---RALSAPACDIIPODEIIPRAHCRLLC--SQOPEVVEK---- 166
Db 454 VPEKLDLSLFMGEGAGLVAHITSRPYVGTVP---LADREWNALLCWDOSQKLLHAWRMFMR 510
QY 167 -ACRSYSHVVLKEVRFFNLQSLYPLLKOPSLNLHIVLVRDPRAVFRSRERTKGLMIDS 225
Db 511 GAWGAHMCVISRE---FLKXVEARLKSSCL-----IAKARR----- 545
QY 226 RIVMGOHEQK-----IKKEDQFYVMQVICOQLEIYKTIQSLPKALQERYLLVRYE 277
Db 546 ----QOHKEKLEAWEVLGKSSDALFRATYLCNARLEPMFSESGL-----RFFLTRGR 595
QY 278 DLRAPVAQTSRMVEFVGLFPLHLOTW---VHNIT--RGKGM 315
Db 596 NNLYGLTNYTEGKRAVTGVQNL-----MSNVVHEVSTKRHKGM 633

Search completed: May 6, 2004, 10:50:27
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:47:22 ; Search time 44 Seconds
(without alignments)
2767.955 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKMKLLFLVSQL.....EQNNLLDLLSTWTVPEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	100.0	386	Q9Y3R3	Q9Y3R3 homo sapien
2	2035	99.9	386	Q8NCG5	Q8NCG5 homo sapien
3	1963	96.3	370	Q8IV46	Q8IV46 homo sapien
4	1515.5	74.4	388	Q9X1I1	Q9X1I1 mus musculus
5	1513.5	74.3	388	Q9WUE5	Q9WUE5 mus musculus
6	1019	50.0	395	Q8GZK3	Q8GZK3 homo sapien
7	1017.5	49.9	395	Q9QUP4	Q9QUP4 mus musculus
8	1008	49.5	411	Q9GZS9	Q9GZS9 homo sapien
9	627	30.8	484	Q9EP78	Q9EP78 mus musculus
10	624	30.6	484	Q99NB0	Q99NB0 mus musculus
11	606.5	29.8	486	Q75667	Q75667 homo sapien
12	605.5	29.7	486	Q9NS84	Q9NS84 homo sapien
13	588.5	29.4	530	Q88276	Q88276 mus musculus
14	595.5	29.2	530	Q80NV3	Q80NV3 mus musculus
15	591.5	29.0	483	Q9UED5	Q9UED5 homo sapien
16	591.5	29.0	530	Q9Y4C5	Q9Y4C5 homo sapien

17	559	27.4	411	11	Q9EQC0
18	549	26.9	411	4	Q43916
19	504.5	24.8	472	11	Q88199
20	482	23.7	479	4	Q75099
21	481.5	23.6	474	11	Q9QZL2
22	423.5	20.8	441	13	Q83403
23	264	13.0	486	5	Q9VMC3
24	215	10.5	120	11	Q9DOK5
25	210	10.3	363	5	Q9VMC4
26	199	9.8	363	5	Q8MZD1
27	175	8.6	119	6	Q95JAB
28	168	8.2	307	16	Q92VG4
29	124.5	6.1	274	16	Q8PRA0
30	118	5.8	368	16	Q8IAP8
31	116.5	5.7	413	16	Q7UP52
32	115.5	5.7	303	16	Q93JEB
33	111	5.6	360	5	Q86R90
34	111	5.4	655	6	Q9BAX0
35	108.5	5.3	1222	4	Q8IZU8
36	107	5.3	1586	4	Q9P2D3
37	100.5	4.9	747	4	Q96QU2
38	100.5	4.9	762	17	Q8OCG7
39	100.5	4.9	775	4	Q9C011
40	100	4.9	593	10	Q8VZE0
41	100	4.9	593	10	Q8ARR2
42	97.5	4.8	285	2	Q8KWX0
43	96.5	4.7	743	4	Q9HOK2
44	96	4.7	2454	3	Q9UV56
45	96	4.7	2454	3	Q9UVP2

ALIGNMENTS

RESULT 1

Q9Y5R3 PRELIMINARY; PRT; 386 AA.

AC Q9Y5R3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
DE sulfotransferase GST-3).
GN GST3.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Tonsil;

RX MEDLINE=92264336; PubMed=10330415;

RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.R.,

RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;

RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";

RL J. Cell Biol. 145:899-910(1999).

RN [2]
RP SEQUENCE FROM N.A.

RC TISSUE=Tonsil;

RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,

RA Rosen S.D.;

RT "Chromosomal Localization and Genomic Organization for the

RT Galactose/N-Acetylglucosamine/N-Acetylglucosamine 6-O-

RT Sulfotransferase Gene Family.";

GL Glycobiology 0:0-0(2001).

RN [3]
RP SEQUENCE FROM N.A.

RC MEDLINE=21332592; PubMed=11439191;

RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellies L.C.,

RA Rabuka D., Hindsaun O., Marth J.D., Lowe J.B., Fukuda M.;

RT "Novel sulfated lymphocyte homing receptors and their control by a

RT corel extension beta1,3-N-acetylglucosaminyltransferase.";

Cell 105:957-969 (2001).
 DR EMBL; AF131235; RAD33015.1; -;
 DR EMBL; AF280088; AK48246.1; -;
 DR EMBL; AF149783; AK48417.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0006928; P:cell motility; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0006955; P:immune response; TAS.
 DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
 DR InterPro; IPR001092; HLH basic.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 KW Lectin; Selectin; Transferase.
 SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;
 Query Match 100.0%; Score 2038; DB 4; Length 386;
 Best Local Similarity 100.0%; Pred. No. 5.5e-183; Indels 0; Gaps 0;
 Matches 386; Conservative 0; Mismatches 0;
 QY 1 MLLPKKMLLLFLVSQMAILAFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
 DB 1 MLLPKKMLLLFLVSQMAILAFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
 QY 61 LFGQHPDVFYLMPEPAHVMWTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFYLMPEPAHVMWTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSIFOWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 DB 121 SSIFOWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 QY 181 FFNLQSLYPLKDPKPSLNHLHVLRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 DB 181 FFNLQSLYPLKDPKPSLNHLHVLRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 QY 241 QPYVWQVICOQLEIYKTIQSLPKALQRYLLVRYEDLARAPVAQTSRMVEFVGLFELP 300
 DB 241 QPYVWQVICOQLEIYKTIQSLPKALQRYLLVRYEDLARAPVAQTSRMVEFVGLFELP 300
 QY 301 HLQTVHNITRKGMDGHAFHTNARDALNVSAQWMSLPYKVSRLQKACGDANLLGYR 360
 DB 301 HLQTVHNITRKGMDGHAFHTNARDALNVSAQWMSLPYKVSRLQKACGDANLLGYR 360
 QY 361 HVRSQEQRNLLDLSTWTVPQIH 386
 DB 361 HVRSQEQRNLLDLSTWTVPQIH 386
 RESULT 2
 Q8NCG5 PRELIMINARY; PRT; 386 AA.
 AC Q8NCG5, 2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein FLJ90265.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isegai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
 RA Masuho Y., Ota T., Okano K., Yoshikawa Y., Aotsuma S., Sasaki N.,
 RA Hattori A., Okumura K., Iwavanagi T., Ninomiya K.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK374746; BAC117.1; -;
 DR Genew; HGNC:1972; CHST4.

GO; GO:0008146; F:sulfotransferase activity; IEA.
 InterPro; IPR000863; Sulfotransferase.
 Pfam; PF00685; Sulfotransfer; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;
 Query Match 99.9%; Score 2035; DB 4; Length 386;
 Best Local Similarity 99.7%; Pred. No. 1.1e-182;
 Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLPKKMLLLFLVSQMAILAFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
 DB 1 MLLPKKMLLLFLVSQMAILAFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
 QY 61 LFGQHPDVFYLMPEPAHVMWTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFYLMPEPAHVMWTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSIFOWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 DB 121 SSIFOWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 QY 181 FFNLQSLYPLKDPKPSLNHLHVLRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 DB 181 FFNLQSLYPLKDPKPSLNHLHVLRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 QY 241 QPYVWQVICOQLEIYKTIQSLPKALQRYLLVRYEDLARAPVAQTSRMVEFVGLFELP 300
 DB 241 QPYVWQVICOQLEIYKTIQSLPKALQRYLLVRYEDLARAPVAQTSRMVEFVGLFELP 300
 QY 301 HLQTVHNITRKGMDGHAFHTNARDALNVSAQWMSLPYKVSRLQKACGDANLLGYR 360
 DB 301 HLQTVHNITRKGMDGHAFHTNARDALNVSAQWMSLPYKVSRLQKACGDANLLGYR 360
 QY 361 HVRSQEQRNLLDLSTWTVPQIH 386
 DB 361 HVRSQEQRNLLDLSTWTVPQIH 386
 RESULT 3
 Q8IV46 PRELIMINARY; PRT; 370 AA.
 AC Q8IV46, 2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC035282; AAH35282.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR GO; GO:0016740; P:transferase activity; IEA.
 DR InterPro; IPR001092; HLH basic.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 KW Transferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;
 Query Match 96.3%; Score 1963; DB 4; Length 370;
 Best Local Similarity 100.0%; Pred. No. 5.9e-176;
 Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 17 MAILAFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQDLFGQHPDVFYLMPEPAW 76

```
Db 1 MAILALPEHMYSHN1SSLSMKAQPERMHV1LSSWRSGSFVQ1FGCHPDVFIEMPAW 60
QY 77 HVMWTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRROSSLFQWENSRLCSAP 136
Db 61 HVMWTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRROSSLFQWENSRLCSAP 120
QY 137 ACDDIIPODEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVRFNLSLYPLKDPSSL 196
Db 121 ACDDIIPODEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVRFNLSLYPLKDPSSL 180
QY 197 NLHIVHLVDRPRAVRSRRTKGLMDSRIVMGHEOKLAKEDQPYVYVMOVICOSOLEI 256
Db 181 NLHIVHLVDRPRAVRSRRTKGLMDSRIVMGHEOKLAKEDQPYVYVMOVICOSOLEI 240
QY 257 YKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFELPHLQVTHWNITRGKMG 315
Db 241 YKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFELPHLQVTHWNITRGKMG 300
QY 317 DHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGYRHRVSEQQRNLLDLL 376
Db 301 DHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGYRHRVSEQQRNLLDLL 360
QY 377 STWTVPEQIH 386
Db 361 STWTVPEQIH 370

RESULT 4
Q9RI11 ID Q9RI11 PRELIMINARY; PRT; 388 AA.
AC Q9RI11;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE L-selectin ligand sulfotransferase.
GN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93361934; PubMed=10435581;
RA Hiraoka N., Petryniak B., Nakayama J., Teubol S., Suzuki M., Yeh J.C.,
RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.";
RL Immunity 11:79-89(1999).
DR EMBL; AF109155; AAD45579.1; -.
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:000529; F:sugar binding; IEA.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR Lectin; Selectin; Transferase.
SQ SEQUENCE 388 AA; 44636 MW; 6D5371AFB6884AEE CRC64;

Query Match 74.4%; Score 1515.5; DB 11; Length 388;
Best Local Similarity 72.8%; Pred. No. 7.1e-134;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MLLPKXMKLLFLVSQMAILALFFHMYSHN1SSLSMKAQPER-MHVLVLSWRSGSFVYG 59
Db 1 MMLKKGRLLMFLGSGVIVVALFIHMSVHR--HLSQREBSRRPVHVLVLSWRSGSFVG 58
QY 60 QLFQGHDPDVFYLMPEAWHVMWTKQSTAMLMHNAVRDLIRAVFLCDMSVFDAYMEPGPR 119
Db 59 QLFQGHDPDVFYLMPEAWHVMWTKQSTAMLMHNAVRDLIRAVFLCDMSVFDAYMEPGPR 118
QY 120 QSSSLFQWENSRLCSAPACDIIIPQDEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEV 179
```

```
Db 119 QSSSLFQWENSRLCSAPACDIIIPQDEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEV 178
QY 180 RFENLQSLYPLKDPSSLNLHIVHLVDRPRAVRSRRTKGLMDSRIVMGHEOKLAKKE 239
Db 179 RFLSQALYPLTDLDSLNLHIVHLVDRPRAVRSRRTKGLMDSRIVMGHEOKLAKKE 238
QY 240 DQPYVYVMOVICOSOLEIYKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFEL 299
Db 239 DQPYVYVMOVICOSOLEIYKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFEL 298
QY 300 PHLQVTHWNITRGKMGDHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGY 359
Db 299 PHLQVTHWNITRGKMGDHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGY 358
QY 360 RHYRSEQQRNLLDLLSTWTVPEQI 385
Db 359 LQVRSQEQGNLSLDLLSSSHILGOV 384

RESULT 5
Q9WUE5 ID Q9WUE5 PRELIMINARY; PRT; 388 AA.
AC Q9WUE5;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin
DE 6/keratan) sulfotransferase 4).
GN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Tonsil;
RX MEDLINE=9264336; PubMed=10330415;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boicelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald C., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Hayashaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF131236; AAD33016.1; -.
DR EMBL; AK009113; BAB26078.1; -.
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
```


RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kaeukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kuchiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli P., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barah G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [4]
RN SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RP Strausberg R.,
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280089; AAG48247.1; -;
DR EMBL; AB040710; BAB13769.1; -;
DR EMBL; AK011202; BAB27465.1; -;
DR EMBL; BC019204; AAI19204.1; -;
DR PIR; JC7350; JC7350.
DR MGD; MGI:1891767; Chat7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0046740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
DR KW Transferase.
SQ SEQUENCE 484 AA; 54766 MW; 9B195537D7AB7193 CRC64;

Query Match 30.8%; Score 627; DB 11; Length 484;
Best Local Similarity 34.4%; Pred. No. 3e-50;
Matches 137; Conservative 65; Mismatches 150; Indels 46; Gaps 7;

QY 30 NISSLSMKAQPERMHVLLSSWRSGSFVGLFGQHPDVFYLMPEAHVMMTFKQSTAWM 89
DB 87 NLSAVEAVTQEKHIYVHATWRTGSSFLGELFNQHPDVFYLYPMHLMQALYPGDAES 146
QY 90 LHMVARDLIRAVFLCDMSVFDAYMEPPGPRQ-----SSLPQWENSRLCSAPACD 139
DB 147 LQALRDMLRSLFRCDFSVRLYAQPGDGERAPDSANLTATMLFRWRTNKVICSPCLCP 206
QY 140 IIPQDE----IIPRAHCRLLCSQPFVEVVEKACRSYSHVWLKEVRFNQLSLYPLLKDPS 195
DB 207 AAPRADVGLVEDKACESTCPVSLRALEAECKYPVVVVKDVLDDLGVLPVLLRDPG 266
QY 140 IIPQDE----IIPRAHCRLLCSQPFVEVVEKACRSYSHVWLKEVRFNQLSLYPLLKDPS 195
DB 207 AAPRADVGLVEDKACESTCPVSLRALEAECKYPVVVVKDVLDDLGVLPVLLRDPG 266
QY 196 LNLHIVHVRDPRVFRSRTKGLMIDS-----RIVM-----GOHE 233
DB 267 LNLKVVQLFRDPRAVHNSRLKSROGLRESIQVLRQGRDHFHVLHAGVDARPGQA 326
QY 234 QKLKEDQPY-----VMQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVQTSR 289
DB 327 RALPSAPRADFFLTSALEVICAWLRDLFTRGAPAWLRRYRLRYEDLVWQQAQLRR 386
QY 290 MYEFVGLGLEFLPHLQTVWNITRGKMG-DHAFHTNARDALNVSAQRWSLPYEKVSRLQK 348
DB 387 LRFSGRLTLAALDAFAFNTRGSAYGADRPFLHSARDAREAVHAWRELRSQEQVRQVET 446
QY 349 ACGDAMNLLGYRHVRSSEQQRNLL-----LDLLSTWTV 381
DB 447 ACAPAMRLLAYPRSGDERDKTVREGTFLETKANNAV 484

RESULT 10
Q99NB0 PRELIMINARY; PRT; 484 AA.
AC Q99NB0;

DT 01-JUN-2001 (T-REMBLrel. 17, Created)
DT 01-JUN-2001 (T-REMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE Chondroitin 6-sulfotransferase-2.
GN CHST7 OR MC6ST-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Kitagawa H., Uyama T., Sugahara K.,
RT "Cloning and Expression of Mouse Chondroitin 6-sulfotransferase-2."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046529; BAB40372.1; -;
DR PIR; JC7350; JC7350.
DR MGD; MGI:1891767; Chat7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
DR KW Transferase.
SQ SEQUENCE 484 AA; 54751 MW; 402C5E1ED185PDF8 CRC64;

Query Match 30.6%; Score 624; DB 11; Length 484;
Best Local Similarity 34.2%; Pred. No. 5.8e-50;
Matches 136; Conservative 66; Mismatches 150; Indels 46; Gaps 7;

QY 30 NISSLSMKAQPERMHVLLSSWRSGSFVGLFGQHPDVFYLMPEAHVMMTFKQSTAWM 89
DB 87 NLSAVEAVTQEKHIYVHATWRTGSSFLGELFNQHPDVFYLYPMHLMQALYPGDAES 146
QY 90 LHMVARDLIRAVFLCDMSVFDAYMEPPGPRQ-----SSLPQWENSRLCSAPACD 139
DB 147 LQALRDMLRSLFRCDFSVRLYAQPGDGERAPDSANLTATMLFRWRTNKVICSPCLCP 206
QY 140 IIPQDE----IIPRAHCRLLCSQPFVEVVEKACRSYSHVWLKEVRFNQLSLYPLLKDPS 195
DB 207 AAPRADVGLVEDKACESTCPVSLRALEAECKYPVVVVKDVLDDLGVLPVLLRDPG 266
QY 196 LNLHIVHVRDPRVFRSRTKGLMIDS-----RIVM-----GOHE 233
DB 267 LNLKVVQLFRDPRAVHNSRLKSROGLRESIQVLRQGRDHFHVLHAGVDARPGQA 326
QY 234 QKLKEDQPY-----VMQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVQTSR 289
DB 327 RALPSAPRADFFLTSALEVICAWLRDLFTRGAPAWLRRYRLRYEDLVWQQAQLRR 386
QY 290 MYEFVGLGLEFLPHLQTVWNITRGKMG-DHAFHTNARDALNVSAQRWSLPYEKVSRLQK 348
DB 387 LRFSGRLTLAALDAFAFNTRGSAYGADRPFLHSARDAREAVHAWRELRSQEQVRQVET 446
QY 349 ACGDAMNLLGYRHVRSSEQQRNLL-----LDLLSTWTV 381
DB 447 ACAPAMRLLAYPRSGDERDKTVREGTFLETKANNAV 484

RESULT 11
Q75667 PRELIMINARY; PRT; 486 AA.
AC Q75667;
DT 01-NOV-1998 (T-REMBLrel. 08, Created)
DT 01-NOV-1998 (T-REMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (Carbohydrate
DE (N-acetylglucosamine 6-O) sulfotransferase 7).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RC TISSUE=Brain;
RA Uchimura K., Muramatsu T.;
RT "Identification and molecular characterization of a cDNA encoding a
RL novel N-acetylglucosamine-6-O-sulfotransferase.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB040711; BAB13770.1; -
DR EMBL; BC045537; AAH45537.1; -
DR FRC; J07351; J07351.
DR Genew; H0NC113817; CHST7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 486 AA; 54266 MW; 31FD1430B3C8E95 CRC64;

Query Match 29.8%; Score 606.5; DB 4; Length 486;
Best Local Similarity 37.0%; Pred. No. 2.6e-48;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERMHVLVLSWRSGSPVQGLFGQHPDVFYLMPEPAHVMVMTFKQSTAWMLHMAVRDLIRA 100
DB 100 EKQHIYHATWRTGSSFLGELFNQHPDVFYLYEPMMHLWQALYPGDAESLQGLRDMRLS 159
QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SSLFQWENSRLCSAPACDIIPQDE---I 146
DB 160 LFRCDPSVLRLYAPPDPAARAPDTANLTAAALFRWRTNKVICSPPLCPGAPARAEEVGL 219
QY 147 IPRAHCRLLCSQQPFVEWEKACSYSHVLEKVRFPNLQSLYPLLLKDPNLNHLVHLVRD 206
DB 220 VEDTACERSCPPVAIRALEAECKYPVWVKDVLRLDGLVFLRLDPGLNLKVQLFRD 279
QY 207 PRAVFRSRETKGDLMTDS-----RIVM-----GOHEQKLKEDQPY 244
DB 280 PRAVHNSRLKSROGLLRRESIQVLRTRQGRDFRHLVLAHGVGAPGQSRALPAAPRADF 339
QY 245 -----VMQVICOQLEIKYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPL 300
DB 340 FLTGALVEICEAWRLDPLFARGAPAWLRRYLRVLYEDLVROPRAQLRLRFSGLRALA 399
QY 301 HLQTVHNITRGKMG-DAHFHTNARDALNVSAWNSLPYKVSRLQKACGDAMNLLGY 359
DB 400 ALDAPALNMTRGAAYGADRPFLHSARDAREAVHWRERLSRQVROVEVACAPAMLLAY 459
QY 360 RHVRSEQE 367
DB 460 --PRSGEE 465

RESULT 12
Q9NS84 Q9NS84 PRELIMINARY; PRT; 486 AA.
AC Q9NS84;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Chondroitin 6-sulfotransferase-2.
GN CHST2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ktagawa H., Fujita M., Ito N., Sugahara K.;
RX MEDLINE=20347214; PubMed=10781596;
RT "Molecular cloning and expression of a novel chondroitin 6-O-
```

```

RT sulfotransferase.";
RL J. Biol. Chem. 275:21075-21080(2000).
DR EMBL; AB037187; BAB03217.1; -
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0008459; F:chondroitin 6-sulfotransferase activity; TAS.
DR GO; GO:0005976; P:polysaccharide metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 486 AA; 54294 MW; B524D1430B3C9075 CRC64;

Query Match 29.7%; Score 605.5; DB 4; Length 486;
Best Local Similarity 37.0%; Pred. No. 3.2e-48;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERMHVLVLSWRSGSPVQGLFGQHPDVFYLMPEPAHVMVMTFKQSTAWMLHMAVRDLIRA 100
DB 100 EKQHIYHATWRTGSSFLGELFNQHPDVFYLYEPMMHLWQALYPGDAESLQGLRDMRLS 159
QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SSLFQWENSRLCSAPACDIIPQDE---I 146
DB 160 LFRCDPSVLRLYAPPDPAARAPDTANLTAAALFRWRTNKVICSPPLCPGAPARAEEVGL 219
QY 147 IPRAHCRLLCSQQPFVEWEKACSYSHVLEKVRFPNLQSLYPLLLKDPNLNHLVHLVRD 206
DB 220 VEDTACERSCPPVAIRALEAECKYPVWVKDVLRLDGLVFLRLDPGLNLKVQLFRD 279
QY 207 PRAVFRSRETKGDLMTDS-----RIVM-----GOHEQKLKEDQPY 244
DB 280 PRAVHNSRLKSROGLLRRESIQVLRTRQGRDFRHLVLAHGVGAPGQSRALPAAPRADF 339
QY 245 -----VMQVICOQLEIKYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPL 300
DB 340 FLTGALVEICEAWRLDPLFARGAPAWLRRYLRVLYEDLVROPRAQLRLRFSGLRALA 399
QY 301 HLQTVHNITRGKMG-DAHFHTNARDALNVSAWNSLPYKVSRLQKACGDAMNLLGY 359
DB 400 ALDAPALNMTRGAAYGADRPFLHSARDAREAVHWRERLSRQVROVEVACAPAMLLAY 459
QY 360 RHVRSEQE 367
DB 460 --PRSGEE 465

RESULT 13
O88276 O88276 PRELIMINARY; PRT; 530 AA.
AC O88276;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN CHST2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
EX MEDLINE=98380482; PubMed=9712885;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RL J. Biol. Chem. 273:22577-22583(1998).
DR EMBL; AB011452; BAA32138.1; -
DR EMBL; AB011452; BAA32139.1; -
DR EMBL; AB011451; BAA32137.1; -
DR MGB; MGI:1891160; Chst2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
```

KW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C363EC CRC64;
Query Match 29.4%; Score 598.5; DB 11; Length 530;
Best Local Similarity 36.8%; Pred. No. 1.6e-47;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;
QY 41 ERMHVLVSSWRSGSFVQGFQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRA 100
DB 163 KQQLVTVFTTWSSGSGFFGELFNQNPVEFFLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 222
QY 101 VFLCMSVFDAYMEPGPRRQS-----SLFQWENSRALCSAPACDIIPQDEIIPRAHCRLL- 155
DB 223 LYRCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVGLVDRVCK 281
QY 156 -CSQQPFVEVEKACRSYSHVVKVRFENLQSLYKVPVDFVAVLAPLKKDPALDKVHLVRDPRVASSR 341
DB 282 KCPQRLARFEECKRYTVIKGVRFVDFVAVLAPLKKDPALDKVHLVRDPRVASSR 341
QY 215 ERTKGLMIDSRIVM-----GQHEOKLKKEDQ----PYV---VMQVTCQS 252
DB 342 IRRHGLIRESLQVVRSDPRHRMPFFLEAAGHKLGAKEGGMGPADYHALGAVEICNS 401
QY 253 QLEIYKTIQSLPKALQERYLLVRYEDLAPAPVAQTSRMVYFVGLFELPHLOTWVHNITRG 312
DB 402 MAKTQTALQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALNMTSG 461
QY 313 KGMGDHAFHTNARDALNVSQAWRWSLPYKVSRLQKACGDAMNLLGYRHVSSEQQRNLL 372
DB 462 SGSSSKPFVVSARNATQAAANWRTALTFOQIKQVEEFCVQPMVGLYERNVSPBEVKDLS 521
QY 373 LDLL 376
DB 522 KTL 525
RESULT 14
Q80WV3 PRELIMINARY; PRT; 530 AA.
ID Q80WV3
AC Q80WV3; 2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to carbohydrate (N-acetylglucosamine-6-O) sulfotransferase
DE 2
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toehyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

[2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RL EMBL; BC051963; AHS1963.1; --
DR GO; GO:0008146; F:transferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;
Query Match 29.2%; Score 595.5; DB 11; Length 530;
Best Local Similarity 36.5%; Pred. No. 3.1e-47;
Matches 133; Conservative 65; Mismatches 137; Indels 29; Gaps 6;
QY 41 ERMHVLVSSWRSGSFVQGFQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRA 100
DB 163 KQQLVTVFTTWSSGSGFFGELFNQNPVEFFLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 222
QY 101 VFLCMSVFDAYMEPGPRRQS-----SLFQWENSRALCSAPACDIIPQDEIIPRAHCRLL- 155
DB 223 LYRCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVGLVDRVCK 281
QY 156 -CSQQPFVEVEKACRSYSHVVKVRFENLQSLYKVPVDFVAVLAPLKKDPALDKVHLVRDPRVASSR 341
DB 282 KCPQRLARFEECKRYTVIKGVRFVDFVAVLAPLKKDPALDKVHLVRDPRVASSR 341
QY 215 ERTKGLMIDSRIVM-----GQHEOKLKKEDQ----PYV---VMQVTCQS 252
DB 342 IRRHGLIRESLQVVRSDPRHRMPFFLEAAGHKLGAKEGGMGPADYHALGAVEICNS 401
QY 253 QLEIYKTIQSLPKALQERYLLVRYEDLAPAPVAQTSRMVYFVGLFELPHLOTWVHNITRG 312
DB 402 MAKTQTALQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALNMTSG 461
QY 313 KGMGDHAFHTNARDALNVSQAWRWSLPYKVSRLQKACGDAMNLLGYRHVSSEQQRNLL 372
DB 462 SGSSSKPFVVSARNATQAAANWRTALTFOQIKQVEEFCVQPMVGLYERNVSPBEVKDLS 521
QY 373 LDLL 376
DB 522 KTL 525
RESULT 15
Q9UED5 PRELIMINARY; PRT; 483 AA.
ID Q9UED5
AC Q9UED5; 2003 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
GN GN6ST.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96391845; PubMed=9722682;
RX Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis x: Molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells."
RL J. Biochem. 124:670-678 (1998).
DR EMBL; AB014679; BAA34265.2; --
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR00863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW Transferase.
 SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;

Query Match 29.0%; Score 591.5; DB 4; Length 483;
 Best Local Similarity 36.3%; Pred. No. 6.5e-47;
 Matches 132; Conservative 66; Mismatches 137; Indels 29; Gaps 6;

QY	41	ERMHVLVLSWSRSGSSFGQLFGQHPDVFYLMERPAWHVWMTFKOSTAWMLHMAVRDLIRA	100
Db	116	KROLVYVTTWRSGSSFFGELFNQNPVEFFLYEPVHWQKLYPGDAVSLOGAARDMLSA	175
QY	101	VFLCDMSVFDAYMEGPRRQS----SLFQWENSRALCSAPACDIPODEIIPRAHCRLL-	155
Db	176	LYRCDLSVFQLYSPAGSGGRNLTTFGFAATNKVCSPLCPAY-RKEVYGLVDDRCK	234
QY	156	-CSQQPFVEVKACSYSHVVLKEVRFENLQSLYELLKDPISNLHLHVLVRDPRAVFRSR	214
Db	235	KCPORLARFEECKRYTIVIGVRFVDVAVLAPLLRDPALDLKVHILVRDPRAVASSR	294
QY	215	ERTKGDLMIDSRIVM-----GQHEQKLKE----DOPYY---VMQVICQS	252
Db	295	ISRHGLIRESIQVVRSDPRHRMPFLEAAGHKLGAKEGVGGPADYHALGAMVEICNS	354
QY	253	QLEIYKTIQSLPKALQERYLLVYEDLARAPVAQTSRMVEFVGLFELPHLOTWVHNTRG	312
Db	355	MAKTLOTALQPPDWJQGHYLVVRYEDLVGDPVKTIRRVYDFVGLLVSPEMEQFALNMTSG	414
QY	313	KGMGDHAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNLLGYRHVRSEQQRNLL	372
Db	415	SGSSSKPFVVSARNATQANANRTALTPOIKQVEEFCVQPMVILGYERVNSPEEVKDLS	474
QY	373	LDLL	376
Db	475	KTLL	478

Search completed: May 6, 2004, 10:51:24
 Job time : 46 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:45:57 ; Search time 58 Seconds
(without alignments)
1880.402 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLPPKMKLLFLVQMAIL.....EORNLLDLLTWTVPQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	100.0	386	2 AAY39918	Aay39918 Human gly
2	2035	99.9	386	4 AAM93309	Aam93309 Human pol
3	2008	98.5	380	5 AAU11274	Aau11274 Human L-s
4	1936	95.0	386	3 AAY79219	Aay79219 Human tra
5	1513.5	74.3	388	2 AAY39919	Aay39919 Mouse gly
6	1205.5	59.2	418	3 AAB41947	Aab41947 Human ORF
7	1028	50.4	418	5 ABB81557	Abb81557 Mouse int
8	1019	50.0	395	4 AAY72640	Aay72640 Human gly
9	1019	50.0	395	5 ABB81554	Abb81554 Human cor
10	1019	50.0	395	5 AAE15438	Aae15438 Human dru
11	1017.5	49.9	395	4 AAY72638	Aay72638 Mouse gly
12	1017.5	49.9	395	5 AAU11275	Aau11275 Murine in
13	1013.5	49.7	395	5 ABB81555	Abb81555 Consensus
14	1008	49.5	390	4 AAY72639	Aay72639 Human gly
15	1008	49.5	390	5 ABB81556	Abb81556 Human int
16	821	40.3	171	5 ABB81560	Abb81560 Human hig
17	605.5	29.7	486	6 ABP56121	Abp56121 Human cho
18	598.5	29.4	483	4 AAY31656	Aay31656 Mouse N-a
19	591.5	29.0	530	4 AAB95367	Aab95367 Human pro
20	587.5	28.8	484	2 AAY31657	Aay31657 Human N-a
21	587.5	28.8	531	5 AAU69414	Aau69414 Lung smal
22	549	26.9	411	2 AAU61100	Aau61100 Keratan s
23	549	26.9	411	5 AAE25356	Aae25356 Human cho
24	549	26.9	411	6 ABU03503	Abu03503 Angiogen
25	519.5	25.5	169	5 ABB81558	Abb81558 Human cor

26	511.5	25.1	169	5 ABB81559	Abb81559 Human int
27	500.5	24.6	458	2 AAW06480	Aaw06480 Chick cho
28	492	24.1	499	6 ABR41139	Abr41139 Human DIT
29	482	23.7	479	2 AAW52863	Aaw52863 Glycosami
30	336	16.5	169	5 ABB81561	Abb81561 Human N-a
31	335.5	16.5	179	5 ABB81562	Abb81562 Human ker
32	293	14.4	174	5 ABB81563	Abb81563 Human cho
33	210	10.3	363	4 ABB64512	Abb64512 Drosophil
34	186.5	9.2	183	4 ABB65852	Abb65852 Mycobacte
35	116	5.7	388	6 AAE32782	Aae32782 Mouse gly
36	109	5.3	1207	4 AAY72643	Aay72643 Human nov
37	109	5.3	1207	4 AAY72641	Aay72641 Human gly
38	108.5	5.3	596	4 AAY72641	Aay72641 Human nov
39	108.5	5.3	1212	6 AAE33541	Aae33541 Human nov
40	108.5	5.3	1222	4 AAY72642	Aay72642 Human gly
41	108.5	5.3	1222	6 ABU11849	Abu11849 Human sec
42	100.5	4.9	775	5 ABB97942	Abb97942 Human pro
43	100	4.9	593	7 ADD30821	Add30821 Plant yie
44	100	4.9	593	7 ADE31489	Ade31489 Plant yie
45	96.5	4.7	743	4 ABUS2986	Abus2986 Human nuc

ALIGNMENTS

RESULT 1
AAY39918
ID AAY39918 standard; protein; 386 AA.

XX AC AAY39918;
XX DT 08-DEC-1999 (first entry)
XX DE Human glycosyl sulfotransferase-3 protein sequence.

XX KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
XX KW selectin binding interaction; inflammation; lymphocyte homing; human;
XX KW secondary lymph organ.

XX OS Homo sapiens.

XX PN WO9949018-A1.

XX PD 30-SEP-1999.

XX PF 26-FEB-1999; 99WO-US004316.

XX PR 20-MAR-1998; 98US-00045284.

XX PR 12-NOV-1998; 98US-00190911.

XX PA (REGC) UNIV CALIFORNIA.

XX PA (SYNT) SYNTX USA INC.

XX PI Bistrup A, Rosen SD, Targemann K, Hemmerich S;

XX DR WPI; 1999-580442/49.

XX DR N-PSDB; AAZ20792.

XX PT Human and murine glycosyl sulfotransferase 3 and related polynucleotides.

XX PS Claim 2; Fig 1; 59pp; English.

XX CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of the
XX CC invention. The nucleic acid sequences, probes and primers derived from
XX CC these, proteins and antibodies are useful in detecting homologues. The
XX CC sequences, antibodies and methods are useful in the diagnosis and
XX CC treatment of diseases associated with selectin binding interactions,
XX CC including conditions associated with or resulting from the homing of
XX CC leukocytes to sites of inflammation and the normal homing of lymphocytes
XX CC to secondary lymph organs

XX SQ Sequence 386 AA;

been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a polypeptide encoded by a full length human cDNA of the invention. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.

```

X      Q      Sequence 386 AA;
      Query Match      99.9%;      Score 2035;      DB 4;      Length 386;
      Best Local Similarity 99.7%;      Pred. No. 2.3e-205;
      Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

1	MLLPKXKLLLFVLSQMAILAFFMYSHNITSSIMKQAPERMHVVLSSWRSGSSFVGQ	60
1	MLLPKXKLLLFVLSQMAILAFFMYSHNITSSIMKQAPERMHVVLSSWRSGSSFVGQ	60
61	LFGQHPDVYLMPEPAWHVMWMTFKQSTAWMLHMAVEDLIRAVFLCDMSVFDAIMEPPGPRQ	120
61	LFGQHPDVYLMPEPAWHVMWMTFKQSTAWMLHMAVEDLIRAVFLCDMSVFDAIMEPPGPRQ	120
121	SSLFOWENSRALCSAPACDIIIPQDEIIIPRAHCELLCSQQPEVVEKACRSYSHVVLKEVR	180
121	SSLFOWENSRALCSAPACDIIIPQDEIIIPRAHCELLCSQQPEVVEKACRSYSHVVLKEVR	180
181	FFNLQSLYPLLKDPSPSLNLIHVLVRDPRAVFRSERTKGDMLIDSRIVMGQHQEQLKKED	240
181	FFNLQSLYPLLKDPSPSLNLIHVLVRDPRAVFRSERTKGDMLIDSRIVMGQHQEQLKKED	240
241	QPYVWQVTCQSQLEIYKTIQSLPKALQERYLLVYEDLARAPVQATSRMYEFVGLFPL	300
241	QPYVWQVTCQSQLEIYKTIQSLPKALQERYLLVYEDLARAPVQATSRMYEFVGLFPL	300
301	HLQTVWHTITRGKMGDGHAFHTNARDALNVSQAWRMSLPEKVSRLQACGQDAMNLLGYR	360
301	HLQTVWHTITRGKMGDGHAFHTNARDALNVSQAWRMSLPEKVSRLQACGQDAMNLLGYR	360
361	HVRSQEQORNLLDLLSTWTVPPQIH	386
361	HVRSQEQORNLLDLLSTWTVPPQIH	386

XX	AAU11274;	
XX	12-MAR-2002	(first entry)
XX	Human L-selectin sulfotransferase-2	(LSST-2) protein.
XX	Human; beta1.3GNT; beta1.3-N-acetylglucosaminyltransferase; MECA-79;	
XX	L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;	
XX	ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;	
XX	allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;	
XX	delayed-type hypersensitivity reaction; hyperplastic thymus; antitumor;	
XX	antiinflammatory; antipsoriatic; antidiabetic; dermatological;	
XX	antiallergic.	
XX	Homo sapiens.	
XX	OS	
XX	PN	WO200185177-A1.
XX	PD	15-NOV-2001.
XX	XX	
XX	10-MAY-2001;	2001WO-US015452.
XX	11-MAY-2000;	2000US-00569320.
XX		
XX		

Query Match	100.0%;	Score 2038;	DB 2;	Length 386;
Best Local Similarity	100.0%;	Pred. No. 1.1e-205;	Indels 0;	Gaps 0;
Matches 386;	Conservative 0;	Mismatches 0;		
1	MLLPKKMKLLLFVLSQMAILALFFHMYSHNISLSMKQAQPERMHVLLVSSWRSSSFVQG	60		
1	MLLPKKMKLLLFVLSQMAILALFFHMYSHNISLSMKQAQPERMHVLLVSSWRSSSFVQG	60		
61	LFQGHDPDFVIMPEAMHVMWMTFFQSTAWMLHMAVRDLIRAVELCDMSVFDAIMEPGRRQ	120		
61	LFQGHDPDFVIMPEAMHVMWMTFFQSTAWMLHMAVRDLIRAVELCDMSVFDAIMEPGRRQ	120		
121	SSLPQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQQPFVVEKACRSYSHVVVLKEVR	180		
121	SSLPQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQQPFVVEKACRSYSHVVVLKEVR	180		
181	FFNLQSLPYLLKDPSLNLHIVHVRPRAVFRSRERTKGDLMDISRIVMQGEQKLKED	240		
181	FFNLQSLPYLLKDPSLNLHIVHVRPRAVFRSRERTKGDLMDISRIVMQGEQKLKED	240		
241	QPYYMQVICSQLEIYKTKQSLPKALOERYILVRVEDLARAPVQTSRMVEFVGLFELP	300		
241	QPYYMQVICSQLEIYKTKQSLPKALOERYILVRVEDLARAPVQTSRMVEFVGLFELP	300		
301	HLQTVWHNITRGKMGDHAFTHTNARDALNYSQAWRWSLPYEKYSRLQKACGDAMNLLGYR	360		
301	HLQTVWHNITRGKMGDHAFTHTNARDALNYSQAWRWSLPYEKYSRLQKACGDAMNLLGYR	360		
361	HYRSEQBORNLLDLLSTWTVPQIHH	386		
361	HYRSEQBORNLLDLLSTWTVPQIHH	386		

RESULT 2
AAM93309
ID AAM93309 standard; protein; 386 AA.
XX
AC AAM93309;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2817.
XX
KW human; full length cDNA; cDNA synthesis; oligo-capping.

XX	Homo sapiens.
XX	
PN	EP1130094-A2.
XX	
PD	05-SEP-2001.
XX	
PF	07-JUL-2000; 2000EP-00114089.
XX	
PR	08-JUL-1999; 99JP-00194486.
PR	11-JAN-2000; 2000JP-00118774.
PR	02-MAY-2000; 2000JP-00183785.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX	
DR	WPI; 2001-524255/58.
DR	N-PSDB; AAK94229.
XX	
PT	830 Primers useful for synthesizing full length cDNA clones and their use
PT	in genetic manipulation.
XX	
PS	Claim 8; SEQ ID NO 2817; 1380pp + Sequence Listing; English.
XX	
CC	The invention relates to primers for synthesising full length cDNA
CC	clones. 830 cDNA molecules encoding a human protein have been isolated
CC	and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have

PA (BURN-) BURNHAM INST.
 XX
 PI Fukuda M, Yeh J, Hiraoka N;
 XX
 DR WPI; 2002-075226/10.
 DR N-PSDB; AAS16947.
 XX
 PT New enzyme, useful for modifying acceptor molecule, comprises an isolated
 PT L-selectin sulfotransferase-2 that directs expression of L-selectin
 PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
 PT GlcNAc 6-sulfotransferase.
 XX
 PS Claim 21; Fig 4; 98pp; English.
 XX
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated beta1,3-N-
 CC acetylglucosaminyltransferase (beta1,3GnT) or an active fragment, where
 CC beta1,3GnT directs expression of a MECA-79 antigen. The invention also
 CC provides a method of treating or preventing an L-selectin-mediated
 CC condition by reducing the expression or activity of a beta1,3GnT that
 CC directs expression of a MECA-79 antigen. This can be done by
 CC administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3GnT,
 CC and/or a beta1,3GnT antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3GnT. The method is useful for treating the expression or
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents human LST-2
 XX
 SQ Sequence 380 AA;
 Query Match 98.5%; Score 2008; DB 5; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.6e-202;
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 MKLLFLVSMATLALFFHMYSHNTSSLSKKAQPERMHVLLSSWRSGSSFVGLFGQHP 66
 Db 1 MKLLFLVSMATLALFFHMYSHNTSSLSKKAQPERMHVLLSSWRSGSSFVGLFGQHP 60
 QY 67 DVFYLMPEAHVWMTFKQSTAWMLHNAVRDLIRAVFLCDMSVFDAYMEFPGRRSSLFQW 125
 Db 61 DVFYLMPEAHVWMTFKQSTAWMLHNAVRDLIRAVFLCDMSVFDAYMEFPGRRSSLFQW 120
 QY 127 ENSRALCSAPACDIIPQDEIIIPAHCRILCSQOPFVVEKACRSYSHVYLKEVRFNLOS 186
 Db 121 ENSRALCSAPACDIIPQDEIIIPAHCRILCSQOPFVVEKACRSYSHVYLKEVRFNLOS 180
 QY 187 LYPCLKDPSLNLHIVHLVDRPRAVFRSRETRTKGDLMDISRIVMGHEQKLKKEQPPYVM 246
 Db 181 LYPCLKDPSLNLHIVHLVDRPRAVFRSRETRTKGDLMDISRIVMGHEQKLKKEQPPYVM 240
 QY 247 QVLCOSOLEYKTKTQSLPQALQERYLLVRYEDLARPAQTSRMVFEVGLBFLPLQVW 306
 Db 241 QVLCOSOLEYKTKTQSLPQALQERYLLVRYEDLARPAQTSRMVFEVGLBFLPLQVW 300
 QY 307 HNTTRGKMGDGHAFHNTNARDALNVSQAWWSLPYKVSRLQKACGDAMNLLGYRHSRSEQ 366
 Db 301 HNTTRGKMGDGHAFHNTNARDALNVSQAWWSLPYKVSRLQKACGDAMNLLGYRHSRSEQ 360
 QY 367 EQRNLLDLLLSTWTVPFQIH 386
 Db 361 EQRNLLDLLLSTWTVPFQIH 380

RESULT 4
 AAY79219
 ID AAY79219 standard; protein; 386 AA.

XX
 AC AAY79219;
 XX
 DT 19-JUN-2000 (first entry)
 XX
 DE Human transferase TRNSFS-11.
 XX
 KW Transferase; TRNSFS-11; human; antitumor; cell proliferation;
 KW gastrointestinal disorder; developmental disorder; genetic disorder;
 KW neurological disorder; reproductive disorder; smooth muscle disorder;
 KW immunological disorder; inflammation; diagnosis; therapy;
 KW N-acetylglucosamine 6-O-sulfotransferase.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Domain 7..23 "transmembrane domain"
 FT Modified-site 30 /note= "potential N-glycosylation"
 FT Modified-site 35 /note= "potential O-phosphorylation"
 FT Modified-site 50 /note= "potential O-phosphorylation"
 FT Modified-site 81 /note= "potential O-phosphorylation"
 FT Modified-site 107 /note= "potential O-phosphorylation"
 FT Modified-site 121 /note= "potential O-phosphorylation"
 FT Modified-site 217 /note= "potential O-phosphorylation"
 FT Modified-site 243 /note= "potential O-phosphorylation"
 FT Modified-site 252 /note= "potential O-phosphorylation"
 FT Modified-site 287 /note= "potential O-phosphorylation"
 FT Modified-site 308 /note= "potential N-glycosylation"
 FT Modified-site 329 /note= "potential N-glycosylation"
 FT Modified-site 364 /note= "potential O-phosphorylation"
 FT Modified-site 380 /note= "potential O-phosphorylation"
 XX
 PN W0200014251-A2.
 XX
 PD 16-MAR-2000.
 XX
 PF 09-SEP-1999; 99WO-US020989.
 XX
 PR 10-SEP-1998; 98US-00150657.
 PR 04-NOV-1998; 98US-00186779.
 PR 11-MAY-1999; 99US-0133642P.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
 PI Hillman JL, Azimzai Y;
 XX
 DR WPI; 2000-256996/22.
 DR N-PSDB; AA294211.
 XX
 PT Human transferase proteins useful for preventing, diagnosing and treating
 PT cancers and developmental, gastrointestinal, genetic, immunological,
 PT neurological, reproductive and smooth muscle disorders.
 XX
 PS Claim 1; Page 90-91; 113pp; English.
 XX
 CC The present sequence is that of human transferase TRNSFS-11, 1 of 15
 CC claimed novel human transferase proteins of the invention (see AAY79209-

23). The sequence was deduced from a cDNA clone (see A294211) isolated from a galbladder library. It shows homology to mouse N- acetylglucosamine 6-C-sulfotransferase. RNF5-11 is expressed in dermal and gastrointestinal tissues, especially those associated with inflammation and cell proliferation. The new human transferases and polypeptides can be used in the diagnosis, prevention and treatment of cancer, developmental disorders, gastrointestinal disorders, genetic disorders, immunological disorders, neurological disorders, reproductive disorders, and smooth muscle disorders. The polypeptides can also be used to raise antibodies, and to screen for agonists and antagonists of transferase activity.

```

SQ      Sequence 386 AA;
Query Match          95.0%;   Score 1936;   DB 3;   Length 386;
Best Local Similarity 95.6%;   Pred. No. 6.4e-195;
Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

```

Qy	1	MLLPKOKULLLFLVSCMAILALFFHMYSHNITSLSMKQAQPERMHVLVLSWRSSGSSFVQG	60
Db	1	MLLPKOKULLLFLVSCMAILALFFHMYSHNITSLSMKQAQPERMHVLVLSWRSSGSSFVQG	60
Qy	61	LFQGHDPVYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYNEPQPRRQ	120
Db	61	LFQGHDPVYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYNEPQPRRQ	120
Qy	121	SSLFQWENSRALCSAPACDIIIPQDESIIPRAHCRLLCSQPFVEVVEKACRSYSHVTLKEVR	180
Db	121	SSLFQWENSRALCSAPACDIIIPQDESSPGLTAGSCAVNSSLKLEKACRSYSHVTLKEVR	180
Qy	181	FFNLQSLYPLLDKPSLNLIHIVHLDPPRAVFRSBERTKGDLMDISRVMGHQEQLKKED	240
Db	181	FFNLQSLYPLLDKPSLNLIHIVHLDPRAPVFRSBERTKGDLMDISRVMGHQEQLKKED	240
Qy	241	QPIYVMQVVCOSOLEIYKTIQSLPALQERYLLVRYEDLARAPVAQTSRMVEFVGLFEFLP	300
Db	241	QPIYVMQVVCOSOLEIYKTIQSLPALQERYLLVRYEDLARAPVAQTSRMVEFVGLFEFLP	300
Qy	301	HLQTTVHNITRKGMDGHAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNLLGYR	360
Db	301	HLQTTVHNITRKGMDGHAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNLLGYR	360
Qy	361	HVSEQEQRNLLDLLSTWTVPQIH	386
Db	361	HVSEQEQRNLLDLLSTWTVPQIH	386

RESULT 5
AAV39919
ID AAV39919 standard: protein: 388 AA.

XX AAY39919:

DT 08-DEC-1999 (first entry)

XX DE Mouse glycosyl sulfotransferase-3 protein sequence.

XX Glycyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.

XX Mus sp.

XX
PN
WO9949018-A1.

XX
PD
30-SEP-1999.

XX
PF 26-FEB-1999; 99WO-US004316.

20-MAR-1998: 98US-00045284.

PR 12-NOV-1998; 9BUS-00190311.
XX

PA (REGC) UNIV CALIFORNIA:

PA (SYNT) SYNTAX USA INC.
XX
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
PI
XX
XX WPI; 1999-580442/49.
DR N-PSDB: AAZ20793.
DR

XX Human and murine glycosyl sulfotransferase 3 and related polynucleotides.
PT
XX
PS Claim 2: Fig 4: 59pp; English.

xx This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of the
cc invention. The nucleic acid sequences, probes and primers derived from
cc these, proteins and antibodies are useful in detecting homologues. The
cc sequences, antibodies and methods are useful in the diagnosis and
cc treatment of diseases associated with selectin binding interactions,
cc including conditions associated with or resulting from the homing of
cc leukocytes to sites of inflammation and the normal homing of lymphocytes
cc to secondary lymph organs

Sequence 388 AA;

Query Match 74.3%; Score 1513.5; DB 2; Length 388;

Best Local Similarity 72.6%; Fred. NO: 2.28-150;
Matches 281. Conservative 49: Mismatches 53:
Indels 3: Gaps 2:

1 MLI.PKKMKI.I.I.FI.VSOMATT.AI.FFHMYSHNTSSLSMKAOPEP-MHVLVLSSWRSGSSFVG 59

1 1007 I PVCPB I MEI CSQVYINVA I EIHMSIVR -- HI.SORFEESRRPVHVI.VLSSWRSGSSFVG 58

CO-OP COMMUNITIES VENTURE CAPITAL TRAVEL.COMSVFDAYMEPGRR 119

118

THE UNIVERSITY OF CHICAGO PRESS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

[illegible][illegible]

DB I 79 RFUSLQATIPUUTIDFSENHVVHVNDKFAVRKNKMLT...AZENNYBDAVZSQA...

QY 240 DQPYVMQVICUSQLEIKYRTIQSLFALQERIKLEVRIEDLAFVAVQICAGIEVCELE

D_b 239 DQPYAMKIIICKSQVDIVKAIQTLPEALQQRYLFLRYEDLVKRAFLAQIINDINFGEDFD 239

QY 300 PHLQTWVHNITRGKGMGDHAFHTNARDALNVSQARWSLPYEKVSRLQKAGJAMNLLGI 359

Db 299 PHLQTWVYNVTRGKGMQHAFTNARNALNVSQAWRWSLPYEKVSQQLQDACGEAMDLLGY 358

QY 360 RHRSEQEQRNLLDLLSTWTVPQI 385

D^b 359 LQVRSQQEQGNLSLDLLSSSHILGQV 384

RESULT 6

AAB41947

ID AAB41947 standard; protein; 418 AA.

AC AAB41947;

XX
DT 09-FEB-2001 (first entry)

XX
DE Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.

XX	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW	antiproliferative; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
KW	vulnerary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antiidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW	antiviral; antibacterial; antifungal; antirheumatic; antichyroid;
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;

XX
PA (REGC) UNIV CALIFORNIA.

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
OS
FN W0200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC76156.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 2599-2600; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 418 AA;
Query Match 59.2%; Score 1205.5; DB 3; Length 418;
Best Local Similarity 61.2%; Pred. No. 71e-118;
Matches 241; Conservative 47; Mismatches 79; Indels 27; Gaps 6;
QY 1 MLLPKXKLLFLVQSMALFHHYSHNIISSLSMKAKQPERMHVLSVSRSSFFVQ 60
41 LLLAQTTCLLIISRP-----GFSPPAGGEDAVHVLVLSVSRSSFFLQ 86
DB 61 LFGQHPDVFYLMPEPAWHVMTTKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
87 LFSQHPDVFYLMPEPAWHVMTTKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 146
QY 121 SSLFQWENGRLACSAPACDIIIPQDEIIIPRAHCRLLCSQPPFEVVEKACRSYHVVLKEVR 180
DB 147 SSLFQWENGRLACSAPACDIIIPQDEIIIPRAHCRLLCSQPPFEVVEKACRSYHVVLKEVR 206
QY 181 FFNLSQLYPLLKDPDSLNLHIVLVRDPAVRFRSRTKGDLMIDRSIVNGQHEQKLKED 240

DB 207 FFNLSQLYPLLKDPDSLNLHIVLVRDPAVRFRSRTKGDLMIDRSIVNGQHEQKLKED 265
QY 241 QPYVMQVICOSEIYK--TIQSLPKAQERYLLVREYEDLARAPVAQTSRMVYFVGLGF 298
DB 266 PHLLIREVCKSHVRIAEATLKP-PFLRGRVLRVFEFLAREPLAEIRALYATGTL 324
QY 299 LPHLQTVHNTIRGKMGD--HAFHTNARDALNVSOAWRWSLPYKVKSRLOKACGDANL 356
DB 325 TPQLEAMHNTHGSGIGKPIEAFTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQ 384
QY 357 LGYHVRSEQORNLILLDL-----STWVPE 383
DB 385 LGYREVSAQOORDLTLDLVLPGRPDHFSWASPD 418
RESULT 7
ID ABB81557 standard; protein; 418 AA.
XX
AC ABB81557;
XX
DT 05-SEP-2002 (first entry)
DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
OS Mus musculus.
XX
FN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-00927602.
XX
PR 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX Fukuda MN, Akama TO;
XX WPI; 2002-507643/54.
DR New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
PS Example 5; Page 24-25; 69pp; English.
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyse sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratectomy. The present sequence
CC represents mouse intestinal N-acetylglucosamine-6-sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
SQ Sequence 418 AA;
Query Match 50.4%; Score 1028; DB 5; Length 418;
Best Local Similarity 56.6%; Pred. No. 3.6e-99;
Matches 215; Conservative 43; Mismatches 112; Indels 10; Gaps 6;
QY 1 MLLPK--KMKLLFLVQSMALFHHYSHNIISSLSMKAKQPERMHVLSVSRSSFFV 58

Db 25 MRLEPSSVMSLLMQTGILV-----LVSQVPS-SPAGLGERVHVLLVLSWRSSESVFV 80
QY 59 GQLFGQHPDVFYLMPEPAHWMTTFKOSTAWMLHMAVRDLIRAVFLCDMSVFPAYNEPGR 118
Db 81 GQLFSQHPDVFYLMPEPAHWMTDLSCGSAPALHMAVRDLIRSVFLCDMDVFPAYL-PWRR 139
QY 119 ROSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQCPPEVVEKACRSYSHVWLKE 178
Db 140 NISDLQWAVSRALCSFPVCEAFARGNISSEVCKPLCATRFGLAQEACSSYSHVWLKE 199
QY 179 VRFENLQSLYPLKDPSSLNLHIVHLVRDPAVRFSRRTKGLMIDSRIMVQGHQKLUK 238
Db 200 VRFENLQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLTNGTWV-E 258
QY 239 EDQPYVMQVTCOSQLEIYKTIQSLPKALQERVLLVRYEDLAPVAQTSRMVVFVGLF 298
Db 259 ADPLRLVNEVCRSHVRIEALHKPFPFQDRVRLVRYEDLADPLTVIRE-YAFTGLQL 318
QY 299 LPHLOTWVHNITRGKMG--DHAFHTNARDALNVQAWRSLPYEKVSRLOKACGDAMNL 356
Db 319 TPQLQTIWNIHTGSGPGARREAFKTSRDALSVQAWRHTLPFAKIRRVQELCGGALQL 378
QY 357 LGRHVRSQEQORNLLDILL 376
Db 379 LGRSVHSELEQRDLSLQLL 398

RESULT 8
AAV72640
ID AAV72640 standard; protein; 395 AA.
XX
AC AAV72640;
XX
DT 02-MAY-2001 (first entry)
DE Human glycosyl sulfotransferase-4beta (GST-4beta).
XX Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1.
XX Homo sapiens.
XX WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US019741.
XX
XX 20-JUL-1999; 99US-0144694P.
XX 13-JUN-2000; 2000US-00593828.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
PI WPI; 2001-138471/14.
XX DR N-PSDB; AAD02697, AAD02700.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications.
XX
XX Claim 3; Fig 4B; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4

CC beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
CC membrane protein useful for inhibiting a binding event between a selectin
CC and a selectin ligand, which comprises contacting the selectin with a non
CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
CC the sulphation activity of GST. GST is also useful in inhibiting a
CC selection mediated binding event. GST is useful in gene therapy to treat
CC disorders such as acute or chronic inflammation, systemic lupus
CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, pernicious
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation
XX
XX Sequence 395 AA;
Query Match 50.08; Score 1019; DB 4; Length 395;
Best Local Similarity 54.5%; Pred. No. 3e-98;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;
QY 2 LLPKMKLLLVLSQMAILALFFHMYSHNISS-SMKAQPERMHLVLSWRSSESVFVQQL 61
Db 14 LLLAQTVLLLVLSGRP-----GPSPPAGGEARVHVLLVLSWRSSESVFVQQL 59
QY 62 FGQHPDVFYLMPEPAHWMTTFKOSTAWMLHMAVRDLIRAVFLCDMSVFPAYNEPGR 121
Db 60 FNQHPDVFYLMPEPAHWMTTLLSQSAAATLHMAVRDLIRSVFLCDMDVFPAYL-PWRNLS 118
QY 122 SLFOWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQCPPEVVEKACRSYSHVWLKEVRF 181
Db 119 DLFQWAVSRALCSFPACSAFPRGAISSEAVCKPLCAFSQFTLAREACRSYSHVWLKEVRF 178
QY 182 FNLSQVLYPLKDPSSLNLHIVHLVRDPAVRFSRRTKGLMIDSRIMVQGHQKLUKEDQ 241
Db 179 FNLSQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLTNGTWV-BADP 237
QY 242 PYYMQVTCOSQLEIYK--TIQSLPKALQERVLLVRYEDLAPVAQTSRMVVFVGLF 299
Db 238 GURVREVCVRSHVRIEALHKP-PFELRGHRLVRYEDLAREPLAETIRALYFTGLSLT 296
QY 300 PHLOTWVHNITRGKMG--DHAFHTNARDALNVQAWRSLPYEKVSRLOKACGDAMNLL 357
Db 297 PQLEAWIHNITGSGPGARREAFKTSRNALNVQAWRHALPFAKIRRVQELCAGALQL 356
QY 358 GYRHVRSQEQORNLLDILL-----STWTVPEQIH 386
Db 357 GYRPVSEDEQRNLDLVLPRGLNGFTWASSTASH 392
RESULT 9
ABB81554
ID ABB81554 standard; protein; 395 AA.
XX
AC ABB81554;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22.
XX
XX Homo sapiens.
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.

```

XX 11-AUG-2000; 2000US-00638211.
XX PR 11-AUG-2000; 2000US-0325773P.
XX XX
XX (FUKU//) FUKUDA M N.
XX (AKAW//) AKAWA T O.
XX PA
XX PI Fukuda MN, Akama TO;
XX XX
XX WPI; 2002-507643/54.
XX DR N-PSDB; AEN89506.
XX XX
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX PT useful for treatment, monitoring and diagnosis of macular corneal
XX PT dystrophy.
XX PT
XX Claim 13; Fig 1A-D; 69pp; English.
XX XX
XX The present sequence represents human corneal N-acetylglucosamine-6-
XX CC sulfotransferase (I), which is able to catalyse sulfation of keratan
XX CC sulfate (KS). Also described is a method for monitoring the effect of
XX CC treatments for macular corneal dystrophy (MCD), and detecting
XX CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX CC ophthalmological activity. (I) can be used to treat or prevent macular
XX CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
XX CC without requiring keratinoplasty or keratotomy
XX XX
XX Query 395 AA;
XX XX
XX Query Match 50.0%; Score 1019; DB 5; Length 395;
XX XX Best Local Similarity 54.5%; Pred. No. 3e-98;
XX XX Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;
XX XX
XX QY 2 LLPKMKLLFLVSQLALFFHMYSHNSSLKMAQPERHVLVLSWRSGSSFVQGL 61
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 14 LLLAQTELLFLVSRP-----GPSDPAGGEARHVLVLSWRSGSSFVQGL 59
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 62 FQGHDPVFYLMEPAMVHWMTFKQSTAWMLHMAVRLTRAVFLCDMSVFDAYNEPGRQOS 121
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 60 FNOHPDVFYLMEPAMVHWMTLSSQSAATLHMAVRLVRSVFLCDMDVFDAYL-PWRRNLS 118
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 122 SLFQWNSRALCSAPACDIIQDRIIPRAHCRLLCSQOPEVVEKACRSYSHVILKEVRF 181
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 119 DLDFQAVSRALCSPPACSAFFRGSAISGEAVCKPLCARQSFLLAREACRSYSHVILKEVRF 178
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 182 FNLQSLYPLLPKPSLNLHIVLVRDPRAVFRSRETKGDLMDISRVLMGQHEQKLKXEDQ 241
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 179 FNLQVLYPLLPSPALNLHIVLVRDPRAVLRSEHQAKALARDNGVILGNTGW-EADP 237
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 242 PYYWQVICOQSOLEYK--TIQSLPKALQERYLLVRVEDLARAPVATQSGMTFVGLLEFL 299
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 238 GLRVREVCRSVRIAEAATLKP-PFPLURGYRLVRFEEDLAREPLAEIALYFTGLSLT 296
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 300 PHLQTVHNITKGMKG--DIAETHNARDALNVSQAWRWSLPYEKYSRLQKACGDAMNLL 357
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 297 PQEAWHNITGSGPGARREAPKTSNNALNVSQAWRHALPFAKIERVQELCAGNLQL 356
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 358 GYHRVSEQQRNLLDLL-----STWTVPEQIH 386
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 357 GYRPVYSEQRNALDLVLPRLNGFTWASSTASH 392
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX XX
XX RESULT 10
XX AAEE15438
XX ID AAEE15438 standard; protein; 395 AA.
XX XX
XX AC AAEE15438;
XX XX
XX DT 12-MAR-2002 (first entry)
XX XX
XX DE Human drug metabolising enzyme (DME)-5.
XX XX
XX KW Human; drug metabolising enzyme; gene therapy; autoimmune disorder;

```

Rosen SD, Lee JK, Hemmerich S;
WPI; 2001-138471/14.
N-PSDB; AAD02696.

New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications.

Claim 3; Fig 2; 128pp; English.

The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
gene is found on chromosome 8E1. GST is a type 2 membrane protein useful
for inhibiting a binding event between a selectin and a selectin ligand,
which comprises contacting the selectin with a non-sulphated selectin
ligand. GST and a small molecular agent that inhibits the sulphation
activity of GST. GST is also useful in inhibiting a selectin mediated
binding event. GST is useful in gene therapy to treat disorders such as
acute or chronic inflammation, systemic lupus erythematosus (SLE),
rheumatoid arthritis, polyarteritis nodosa, polymyositis,
dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
disease, adrenalitis, hypoparathyroidism, pernicious anaemia,
demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,
myocarditis, regional enteritis, adult respiratory distress syndrome,
infantile eczema, psoriasis, lichen planus, allergic rhinitis, bronchial
asthma, hypersensitivity, rheumatic fever and tissue rejection during
transplantation

Sequence 395 AA;

Query Match 49.8%; Score 1017.5; DB 4; Length 395;
Best Local Similarity 56.4%; Pred. No. 4.3e-98;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

QY 1 MLPLK--KMKLLFLVSQMAILALFPHMYSHNISSLMKAQPERGHVVLVLSWSSGSFV 58
DB 1 MELPFGSTVMSLSLAYQTGLVF---LVSROVPS--SPAGLGERVHVVLVLSWSSGSFV 56
QY 59 GOLFGQPDVFLYMEPAHWYMMTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEQPR 118
DB 57 GOLFSQPDVFLYMEPAHWYMMTSLQSGAPALFMAVRDLIRSVFLCDMDVDAYL-PWR 115
QY 119 KSSLFOWNSRALCSAPACDIIPQEIIPRAHCRLLCSQPFVEVEKACSYSHVVLKE 178
DB 116 NISLFWAVNSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKE 175
QY 179 VRFFNLQSLPLKDPDLSNLHIVHLVRDPRAVFRSRETKGLMDISIVNGQHEQKLK 238
DB 176 VRFFNLQVLPILSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGILGTNGTWV-E 234
QY 239 EDQPYVMQVICSQLEIYK-TTQSLPKALQRYLLVRYEDLARAPVAQTSRMVFFGLE 297
DB 235 ADPLRLVYNEVCRSHVRAEAAHLKPPPLQDQRYLRYEDLARDPPLTVIRELYAFTGLG 294
QY 298 FLPLHQTWVHNITRGKMG--DIAFPTNARDALNYSQAWNSLPYKVSRLQKACGDMN 355
DB 295 LTPQLQTWNIHTHSGSPGARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCGGALQ 354
QY 356 LLGHRVRSQEQRNLLDLL 376
DB 355 LLGVRSHSELEQRDLSDLL 375

RESULT 12
AAU11275
ID AAU11275 standard; protein; 395 AA.
XX AC AAU11275;
DT 12-MAR-2002 (first entry)
XX DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.

XX Mouse; beta1.3Gnt; beta1.3-N-acetylglucosaminyltransferase; MECA-79;
KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antitumor;
KW antinflammatory; antipsoriatic; antidiabetic; dermatological;
KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; 1-GlcNAc6ST.
XX Mus musculus.
XX WO200185177-A1.
XX 15-NOV-2001.
XX 10-MAY-2001; 2001WO-US015452.
XX 11-MAY-2000; 2000US-00569320.
XX (BURN-) BURNHAM INST.
XX Fukuda M, Yeh J, Hiraoka N;
XX WPI; 2002-075226/10.
XX N-PSDB; AAS16948.
XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
PT L-selectin sulfotransferase-2 that directs expression of L-selectin
PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
PT GlcNAc 6-sulfotransferase.
XX Claim 28; Fig 10; 98pp; English.
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated beta1.3-N-
CC acetylglucosaminyltransferase (beta1.3Gnt) or an active fragment, where
CC beta1.3Gnt directs expression of a MECA-79 antigen. The invention also
CC provides a method of treating or preventing an L-selectin-mediated
CC condition by reducing the expression or activity of a beta1.3Gnt that
CC directs expression of a MECA-79 antigen. This can be done by
CC administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1.3Gnt,
CC and/or a beta1.3Gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LSST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1.3Gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents mouse 1-GlcNAc6ST
XX SQ Sequence 395 AA;
Query Match 49.9%; Score 1017.5; DB 5; Length 395;
Best Local Similarity 56.4%; Pred. No. 4.3e-98;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;
QY 1 MLPLK--KMKLLPLVSCMAILLAFHMYSHNLSLSMKQAQPERMHVILSSWRSSSV 58
Db 1 MRLPRFSSTVMSLLMVTQTGLV--LVSRQVPS-SPAGLGERVHVILSSWRSSSV 56
QY 59 GQLFGQHPDVFYLNPEAHVMMTPKQSTAMKLMHVRDLIRAVFLCDMSVFDAYMPPGPR 118
Db 57 GQLFSQHPDVFYLNPEAHVMMTDLISQSPALHVAVEDLIRSVFLCDMDVFDAYL-PWR 115
QY 119 RQSSLFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQOPPEVVEKACRSYHVVLKE 178
Db 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYHVVLKE 175
QY 179 VRFFNLQSLYLLKDPISLNLHIVLVRDPAVFRSRRRTKGDLMIDSRVGMQGEQKLK 238

Db 176 VRFFNLQSLYLLKDPISLNLHIVLVRDPAVFRSRRRTKALARDNGIVLGTNGTWV-E 234
QY 239 EQOPYVWMOVICQSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFFVGL 297
Db 235 APRLRVNVNVCVRSHVIAEALHKPPPLQDRVRLVRYEDLARDPPLVIRELYAFTGLG 294
QY 298 FLPHTQVHNITRGKMG--DHAPHYARDALNVSOAWRWSLPYKYSRLQKACGDAMN 355
Db 295 LTPQLQCTWHNITHGSGPGARREAFKTTSRDALSVSQAWRHTLTPFAKIRRVQELCGGALQ 354
QY 356 LLGYRHVRSEQRNLLDLL 376
Db 355 LLGYRSVHSELRDLSLDLL 375
RESULT 13
ABB81555
ID ABB81555 standard; protein; 395 AA.
XX ABB81555;
XX AC ABB81555;
XX DT DT
XX 05-SEP-2002 (first entry)
XX Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX Homo sapiens.
OS Mus musculus.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 10 /label= Ala, Thr, Val
FT Misc-difference 13 /label= Ala, Val, Ser
FT Misc-difference 20 /label= Phe, Cys, Gly
FT Misc-difference 39 /label= Ala, Asp, Glu
FT Misc-difference 96 /label= Val, Met, Ile
FT Misc-difference 142 /label= Ala, Thr, Asn
FT Misc-difference 147 /label= Ala, Asp, Glu
FT Misc-difference 159 /label= Thr, Ser, Gly
FT Misc-difference 238 /label= Gly, His, Arg
FT Misc-difference 294 /label= Ser, Thr, Gly
FT Misc-difference 371 /label= Ala, Thr, Ser
FT Misc-difference 380 /label= Leu, Pro, Met
FT Misc-difference 382 /label= Gly, His, Ser
FT Misc-difference 384 /label= Thr, Ser, Lys
FT Misc-difference 390 /label= Ala, Glu
FT Misc-difference 391 /label= Ser, Lys
FT Misc-difference 392 /label= His, Gln
FT Misc-difference 394 /label= Arg, Glu
FT Misc-difference 395 /label= Asn, Ser
FT

XX US2002061562-A1.
XX 23-MAY-2002.
XX 09-AUG-2001; 2001US-00927602.
XX 11-AUG-2000; 2000US-00638211.
XX 11-AUG-2000; 2000US-0325773P.
XX (FUKU// FUKUDA M N.
XX (AKAM// AKAMA T O.
XX FUKUDA MN, Akama TO;
XX WPI; 2002-507643/54.
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX Example 5; Fig 2A-B; 69pp; English.
XX The present invention describes human corneal N-acetylglucosamine-6-
XX sulfotransferase (I), which is able to catalyze sulfation of keratan
XX sulfate (KS). Also described is a method for monitoring the effect of
XX treatments for macular corneal dystrophy (MCD), and detecting
XX susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX ophthalmological activity. (II) can be used to treat or prevent macular
XX corneal dystrophy types I or II. (I) makes possible treatment of MCD
XX without requiring keratoplasty or keratectomy. The present sequence
XX represents a consensus N-acetylglucosamine-6-sulfotransferase which is
XX given in the exemplification of the present invention
XX SQ Sequence 395 AA;
Query Match 49.7%; Score 1013.5; DB 5; Length 395;
Best Local Similarity 55.9%; Pred. No. 1.1e-97;
Matches 212; Conservative 45; Mismatches 101; Indels 21; Gaps 6;
QY 2 LLPKPKMLLLFLVQMAIALFFHMYSHNISLSNKAQPERHVLVLSWRSWSSGSGFVQOL 61
DB 14 LLLAQTXLLFLVSRP-----GPPSSPAGGEXRFXVFLVLSWRSWSSGSGFVQOL 59
QY 62 FQCHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEGPRRQS 121
DB 60 FQCHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEGPRRQS 118
QY 122 SLFQWENSRALCSAPACDIIPQEIIPRAHCRLLCSQPFVEYKACRSYSHVVLKEVRF 181
DB 119 DLFQWAVSRALCSPPACSAFPRGXISSEKVCPLCARQPFXLAREACRSYSHVVLKEVRF 178
QY 182 FNLQSLYPLKDPNLSNLHIVLVRDPAVFRSRETKGDLMDSDRIWQHEQKLKEDQ 241
DB 179 FNLQVLYPLSDPALNLRVHLVRDPAVFRSRETKALARDNGIVLGTNGTWV-EADP 237
QY 242 PYYWMOVICQSOLEYK---TIQSLPKALQERVLLVRVEDLAPVAQTSRYVEFVGLLEFL 299
DB 238 XLRVVREVCESHVIRAEATLKP-PPFLGRVLRVFEEDLAREPLAEIRALYAFYGLXLT 296
QY 300 PHLQTVHNTGKNG--DHAFHNTNARDALNVSAWRNSLPYKVSRLQACGDAMNLL 357
DB 297 PQLAEIHNITHGSGPGARREAFKTSRRNALNVSAWRHALPFKIRRVQELCAGALQL 356
QY 358 GYRHRVASEQQRNLLDL 376
DB 357 GYRFPVSEDEQRDLXLDIV 375
RESULT 14
ID AAY72639
XX AAY72639 standard; protein; 390 AA.

AC AAY72639;
XX 02-MAY-2001 (first entry)
XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1.
XX Homo sapiens.
XX WO200106015-A1.
XX 25-JAN-2001.
XX 19-JUL-2000; 2000WO-US019741.
XX 20-JUL-1999; 99US-0144694P.
XX 13-JUN-2000; 2000US-00593828.
XX (REGC) UNIV CALIFORNIA.
XX Rosen SB, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
XX N-PSDB; AAD02697, AAD02698, AAD02699.
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications.
XX Claim 3; Fig 1; 128pp; English.
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
XX membrane protein useful for inhibiting a binding event between a selectin
XX and a selectin ligand, which comprises contacting the selectin with a non
XX -sulphated selectin ligand, GST and a small molecular agent that inhibits
XX the sulphation activity of GST. GST is also useful in inhibiting a
XX selectin mediated binding event. GST is useful in gene therapy to treat
XX disorders such as acute or chronic inflammation, systemic lupus
XX erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
XX polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation. Note: the present sequence is also shown in
XX sequence listing (page no: 56) but lacks four nucleotides at its 3' end
XX SQ Sequence 390 AA;
Query Match 49.5%; Score 1008; DB 4; Length 390;
Best Local Similarity 52.3%; Pred. No. 4.2e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 1 MLLPKPKMLLLFLVQMAIALFFHMYSHNISLSNKAQPERHVLVLSWRSWSSGSGFVQOL 60
DB 14 LLLAQTXLLFLVSRP-----GPPSSPAGGEXRFXVFLVLSWRSWSSGSGFVQOL 59
QY 61 LFCQCHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEGPRRQS 120
DB 60 LFCQCHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEGPRRQS 118

```
QY 121 SSLFOWNSRALCSAPACDIIIPQDEIIIPRAHCRLLCSQOPPEWVEKACRSYSHVVLKEVR 180
Db 119 SAFFNWTATSRALCSPACSAFPRGTISKQDVCKTICTQPSLAREACRSYSHVVLKEVR 178
QY 181 FNNLSLYPLLDKPSLNLIHIVLDRPRAVFSRERTKGDLMDSRIVNGQHEOKLKKED 240
Db 179 FNNLSLYPLLDKPSLNLIHIVLDRPRAVFSRERTKGDLMDSRIVNGQHEOKLKKED 237
QY 241 QPYVYMOVICQSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAOTSRYMYEFGLEF 298
Db 238 PHLRLEIVCRSHVRIAEANTLKP--PPFLRGYRLVRFEDLAREPLAIRALYFTGLTL 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVQAWRWSLPYKVSRLQKACGDAMNL 356
Db 297 TPQLEAWIHNITHGSGIGKPIEAFTSSRNARNVQAWRHALPFTKILRVQEVCAQALQ 356
QY 357 LGYRVYSADQQRDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGYRVYSADQQRDLTLDLVLPRGPDHFSWASPD 390

RESULT 15
ID ABB81556
XX ABB81556 standard; protein; 390 AA.
AC ABB81556;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc-6ST;
KW cornearl; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Homo sapiens.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-00927602.
XX
PR 11-AUG-2000; 2000US-00638211.
XX
PR 11-AUG-2000; 2000US-0325773P.
XX
PA (FUKU//) FUKUDA M N.
XX
PA (AKAM//) AKAMA T O.
XX
PI Fukuda MN, Akama TO;
XX
PI WPI; 2002-507643/54.
XX
PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
PS Example 5; Fig 2A-B; 69pp; English.
XX
CC The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratinoplasty or keratectomy. The present sequence
CC represents human intestinal N-acetylglucosamine-6-sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
SQ Sequence 390 AA;
```

```
Query Match 49.5%; Score 1008; DB 5; Length 390;
Best Local Similarity 52.3%; Pred.No. 4.2e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 1 MLLPKKKKLLFLVSSQMAILALFFHYSHNSSLKMAQPERMHVILVSSWSGSGSFVQ 60
Db 14 LLQAOTTCULLFIISRP-----CPSSPAGGEDRVHVLVSSWSGSGSFVQ 59
QY 61 LFGQHPDVLYLMEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRKQ 120
Db 60 LFSOHPDVLYLMEPAWHVMTFKQSSAATLHMAVRDLIRAVFLCDMDVFDAYM-PCSRNL 118
QY 121 SSLFOWNSRALCSAPACDIIIPQDEIIIPRAHCRLLCSQOPPEWVEKACRSYSHVVLKEVR 180
Db 119 SAFFNWTATSRALCSPACSAFPRGTISKQDVCKTICTQPSLAREACRSYSHVVLKEVR 178
QY 181 FNNLSLYPLLDKPSLNLIHIVLDRPRAVFSRERTKGDLMDSRIVNGQHEOKLKKED 240
Db 179 FNNLSLYPLLDKPSLNLIHIVLDRPRAVFSRERTKGDLMDSRIVNGQHEOKLKKED 237
QY 241 QPYVYMOVICQSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAOTSRYMYEFGLEF 298
Db 238 PHLRLEIVCRSHVRIAEANTLKP--PPFLRGYRLVRFEDLAREPLAIRALYFTGLTL 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVQAWRWSLPYKVSRLQKACGDAMNL 356
Db 297 TPQLEAWIHNITHGSGIGKPIEAFTSSRNARNVQAWRHALPFTKILRVQEVCAQALQ 356
QY 357 LGYRVYSADQQRDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGYRVYSADQQRDLTLDLVLPRGPDHFSWASPD 390
```

Search completed: May 6, 2004, 10:49:57
Job time : 60 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:51:28 ; Search time 48 Seconds
(without alignments)
2232.099 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKMKLLFLVSQMAIL.....EORNLLDLLSTWTVEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/1/pubaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/1/pubaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/prodata/1/pubaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/prodata/1/pubaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/1/pubaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/prodata/1/pubaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/prodata/1/pubaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/prodata/1/pubaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/prodata/1/pubaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/prodata/1/pubaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/prodata/1/pubaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/prodata/1/pubaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/prodata/1/pubaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/prodata/1/pubaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/prodata/1/pubaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/prodata/1/pubaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/prodata/1/pubaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/prodata/1/pubaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2038	100.0	386	9 US-09-816-825-2	Sequence 2, Appli
2	2038	100.0	386	13 US-10-007-262-1	Sequence 1, Appli
3	1936	95.0	386	14 US-10-427-631-11	Sequence 11, Appl
4	1028	50.4	418	9 US-09-927-602-5	Sequence 5, Appli
5	1019	50.0	395	9 US-09-927-602-2	Sequence 2, Appli
6	1019	50.0	395	12 US-10-258-080-5	Sequence 5, Appli
7	1013.5	49.7	395	9 US-09-927-602-3	Sequence 3, Appli
8	1008	49.5	390	9 US-09-927-602-4	Sequence 4, Appli
9	821	40.3	171	9 US-09-927-602-8	Sequence 8, Appli
10	598.5	29.4	483	14 US-10-212-933-2	Sequence 2, Appli
11	587.5	28.8	484	14 US-10-212-933-4	Sequence 4, Appli
12	587.5	28.8	531	9 US-09-833-790-255	Sequence 255, Appl
13	549	26.9	411	12 US-10-211-462-97	Sequence 97, Appl
14	549	26.9	411	14 US-10-021-660-128	Sequence 128, Appl
15	519.5	25.5	169	9 US-09-927-602-6	Sequence 6, Appli

16	511.5	25.1	169	9	US-09-927-602-7	Sequence 7, Appli
17	488	24.8	481	12	US-10-087-192-123	Sequence 123, App
18	488	23.9	479	12	US-10-087-192-126	Sequence 126, App
19	336	16.5	169	9	US-09-927-602-9	Sequence 9, Appli
20	335.5	16.5	179	9	US-09-927-602-10	Sequence 10, Appli
21	293	14.4	174	9	US-09-927-602-11	Sequence 11, Appli
22	116	5.7	387	14	US-10-126-279-21	Sequence 21, Appli
23	116	5.7	387	14	US-10-286-608-21	Sequence 21, Appli
24	100.5	4.9	668	10	US-09-811-469-6	Sequence 6, Appli
25	100.5	4.9	668	14	US-10-370-659-6	Sequence 39283, A
26	100.5	4.9	807	12	US-10-425-114-39283	Sequence 854, App
27	100	4.9	593	15	US-10-225-066A-854	Sequence 56, Appli
28	100	4.9	593	15	US-10-225-067-56	Sequence 330, App
29	100	4.9	593	15	US-10-374-780A-330	Sequence 149380,
30	98	4.8	629	12	US-10-424-553-149380	Sequence 16, Appli
31	96.5	4.7	791	16	US-10-380-727-16	Sequence 38, Appli
32	95.5	4.7	568	12	US-10-042-865-38	Sequence 132, App
33	95.5	4.7	568	15	US-10-262-445-132	Sequence 380, App
34	93	4.6	348	15	US-10-259-194A-380	Sequence 11, Appli
35	92	4.5	359	14	US-10-411-976-11	Sequence 1075, Ap
36	91	4.5	1847	15	US-10-369-493-1075	Sequence 2, Appli
37	90	4.4	393	14	US-10-126-279-2	Sequence 2, Appli
38	89.5	4.4	4273	15	US-10-286-606-2	Sequence 23144, A
39	89.5	4.4	4273	15	US-10-369-493-23144	Sequence 20, Appli
40	89	4.4	320	9	US-09-854-122-20	Sequence 76479, A
41	89	4.4	1140	12	US-10-282-122A-76479	Sequence 15, Appli
42	89	4.4	1968	14	US-10-223-070-15	Sequence 62635, A
43	88	4.3	745	12	US-10-425-114-62635	Sequence 1903, Ap
44	88	4.3	2209	9	US-09-902-941-1903	Sequence 1903, Ap
45	88	4.3	2209	9	US-09-849-826-1903	

ALIGNMENTS

RESULT 1
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Query Match	100.0%;	Score	2038;	DB	9;	Length	386;
Best Local Similarity	100.0%;	Pred. No.	4.6e-202;				
Matches	386;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MLLPKMKLLFLVSQMAILALFFHMYSHN1SSLSMKAQPERMHVLSWSRSGSFVQG	60				
Db	1	MLLPKMKLLFLVSQMAILALFFHMYSHN1SSLSMKAQPERMHVLSWSRSGSFVQG	60				
QY	61	LFQGPDPVFLMEPAHWVMTFKQSTAWMLHVAVDLIRAVFLCMGVDFDAYMEGPRRQ	120				
Db	61	LFQGPDPVFLMEPAHWVMTFKQSTAWMLHVAVDLIRAVFLCMGVDFDAYMEGPRRQ	120				
QY	121	SSLFOWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOPEVVEKACRSYSHVVLKEVR	180				
Db	121	SSLFOWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOPEVVEKACRSYSHVVLKEVR	180				

QY 181 FFLQSLYPLKDPKPSLNHLHVLVRDPRVFRSRETKGDLMDSRVNGQHEQKLKED 240
Db 181 FFLQSLYPLKDPKPSLNHLHVLVRDPRVFRSRETKGDLMDSRVNGQHEQKLKED 240
QY 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
Db 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
QY 301 HLQVTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
Db 301 HLQVTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
QY 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

RESULT 2

US-10-007-262-1
; Sequence 1, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: COSLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yaida
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-007-262-1

Query Match 100.0%; Score 2038; DB 13; Length 386;
Best Local Similarity 100.0%; Pred. No. 4.6e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAPERMHVILVLSWRSFVQ 60
Db 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAPERMHVILVLSWRSFVQ 60
QY 61 LFGQHPDVFLMEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
Db 61 LFGQHPDVFLMEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
QY 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
Db 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
QY 181 FFLQSLYPLKDPKPSLNHLHVLVRDPRVFRSRETKGDLMDSRVNGQHEQKLKED 240
Db 181 FFLQSLYPLKDPKPSLNHLHVLVRDPRVFRSRETKGDLMDSRVNGQHEQKLKED 240
QY 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
Db 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
QY 301 HLQVTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
Db 301 HLQVTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
QY 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

RESULT 4

US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:

RESULT 3
US-10-427-631-11
; Sequence 11, Application US/10427631
; Publication No. US20030175923A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: COSLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yaida
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-427-631-11

Query Match 95.0%; Score 1936; DB 14; Length 386;
Best Local Similarity 95.6%; Pred. No. 1.7e-191;
Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAPERMHVILVLSWRSFVQ 60
Db 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAPERMHVILVLSWRSFVQ 60
QY 61 LFGQHPDVFLMEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
Db 61 LFGQHPDVFLMEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
QY 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
Db 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
QY 181 FFLQSLYPLKDPKPSLNHLHVLVRDPRVFRSRETKGDLMDSRVNGQHEQKLKED 240
Db 181 FFLQSLYPLKDPKPSLNHLHVLVRDPRVFRSRETKGDLMDSRVNGQHEQKLKED 240
QY 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
Db 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
QY 301 HLQVTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
Db 301 HLQVTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
QY 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

```

; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US 09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

```

```

Query Match 50.4%; Score 1028; DB 9; Length 418;
Best Local Similarity 56.6%; Pred. No. 2e-97;
Matches 215; Conservative 43; Mismatches 112; Indels 10; Gaps 6;

QY 1 MLLPK--KMKLLFLVSQMAILALPHMYSHNLSMKQAQPERMHVLSVSRSSGFV 58
DB 25 MLLPSSVMSLLMVQGIIVF---LVSROVPS-SPAGLGERVHVLSVSRSSGFV 80
QY 59 GQFGQHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGR 118
DB 81 GQFSGHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYL-PWR 139
QY 119 RQSSLPQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOQFVVEVKACRSYSHVLE 178
DB 140 NISDLQWAVSRALCSAPACDIIPQDEIIPRAHCRLLCSQOQFVVEVKACRSYSHVLE 199
QY 179 VRFNQLQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLK 238
DB 200 VRFNQLQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLK 258
QY 239 EDPQVYVMOVICQSLQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 298
DB 259 ADPLRVNVECRSHVRIAEATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYFTGLSLT 318
QY 299 LPHQWVNTIRGKMG--DHAFHTNARDALNVSOAWRSLPYEKVSRLOKACGDAMNLL 356
DB 319 TPQLQWVNTIRGKMG--DHAFHTNARDALNVSOAWRSLPYEKVSRLOKACGDAMNLL 378
QY 357 LGYRVSRQEQRLNLLDLL-----STWTVPEQIH 386
DB 379 LGYRVSRQEQRLNLLDLL-----STWTVPEQIH 398

```

```

RESULT 5
US-09-927-602-2
; Sequence 2, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Akama, Tomoya O.
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-2

```

Query Match 50.0%; Score 1019; DB 9; Length 395;

```

Best Local Similarity 54.5%; Pred. No. 1.6e-96;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LLPKMKLLFLVSQMAILALPHMYSHNLSMKQAQPERMHVLSVSRSSGFVQGL 61
DB 14 LLLAQTFLLFLVSRP-----GPSSPAGGEARVHVLSVSRSSGFVQGL 59
QY 62 FGQHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGRROS 121
DB 60 FNQHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYL-PWRNLS 118
QY 122 SLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOQFVVEVKACRSYSHVLEKVRP 181
DB 119 DLQWAVSRALCSAPACDIIPQDEIIPRAHCRLLCSQOQFVVEVKACRSYSHVLEKVRP 178
QY 182 FNQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLKDEQ 241
DB 179 FNQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLKDEQ 237
QY 242 PYVMOVICQSLQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 299
DB 238 GLRVVRECRSHVRIAEATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYFTGLSLT 296
QY 300 PHQWVNTIRGKMG--DHAFHTNARDALNVSOAWRSLPYEKVSRLOKACGDAMNLL 357
DB 297 PQLQWVNTIRGKMG--DHAFHTNARDALNVSOAWRSLPYEKVSRLOKACGDAMNLL 356
QY 358 GYRVSRQEQRLNLLDLL-----STWTVPEQIH 386
DB 357 GYRVSRQEQRLNLLDLL-----STWTVPEQIH 392

```

```

RESULT 6
US-10-258-080-5
; Sequence 5, Application US/10258080
; Publication No. US20040029125A1
; GENERAL INFORMATION:
; APPLICANT: Incyte Genomics, Inc.
; APPLICANT: POLICKY, Jennifer L.
; APPLICANT: HAPALIA, April J.A.
; APPLICANT: BURFORD, Neil
; APPLICANT: RING, Huijun Z.
; APPLICANT: LAL, Preeti G.
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: YAO, Monique G.
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: DAS, Debopriya
; APPLICANT: SANJANWALA, Madhusudan M.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: REDDY, Roopa M.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: RAMKUNAR, Jayalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: AU-YOUNG, Janice K.
; TITLE OF INVENTION: DRUG METABOLIZING ENZYMES
; FILE REFERENCE: PI-0070 USN
; CURRENT APPLICATION NUMBER: US/10/258,080
; CURRENT FILING DATE: 2002-10-15
; PRIOR FILING DATE: 2000-05-11
; PRIOR FILING DATE: 2000-05-11
; PRIOR FILING DATE: 2000-05-11
; PRIOR FILING DATE: 2000-05-05
; PRIOR FILING DATE: 2000-04-29
; PRIOR FILING DATE: 2000-04-29
; PRIOR FILING DATE: 2000-04-19
; PRIOR FILING DATE: 2001-04-12
; PRIOR FILING DATE: 2000-04-13

```

```

; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040029125A1 7472777CD1
US-10-258-080-5

Query Match
Best Local Similarity 50.0%; Score 1019; DB 12; Length 395;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LLPKMKLLFLVVSQMAILALFFHMYSHNISLSMKAQPERMHVLLVLSWSSGSGSFVQQL 61
DB 14 LLLAQTFLLFLVSRP-----GSSSPAGGEARVHVLLVLSWSSGSGSFVQQL 59
QY 62 FQGHDPDVFYLMPEPAWVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 121
DB 60 FQGHDPDVFYLMPEPAWVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 118
QY 122 SLFQWNSRALCSAPACDIIPDEIIPRAHCELLCSQOPFEVVEKACRSYSHVVLKEVRF 181
DB 119 DLFOVAVSRALCSPPACSAFPRGXISSEKVCPLCARQPFKLAREACRSYSHVVLKEVRF 178
QY 182 FNLQSLYPLLKDPSPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 241
DB 179 FNLQVLYPLSDPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 237
QY 242 PYYMQVICOSQLEIYK--TIOSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLLEFL 299
DB 238 GLRVVREVCRSVRVIAEAATLKP-PFFLRGRYLRVRFEDLAREPLAEIRALYAFGLXLT 296
QY 300 PHLOTWVHNITRGKMG--DHAFHTNARDALNVSAWRSLPYEKVSRLOKACGDAMNLL 357
DB 297 PQLKAWIHNITGSGPGARREAFKTSRNALNVSAWRHALPFAKIRVQELCAGALQLL 356
QY 358 GYRHRVSEQQRNLLDLL-----STWVPEQIH 386
DB 357 GYRVPVSEDEQRLDLVLPRLGNGFTWASSTASH 392

RESULT 7
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(395)
; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match
Best Local Similarity 49.7%; Score 1013.5; DB 9; Length 395;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 1 MLLPKMKLLFLVVSQMAILALFFHMYSHNISLSMKAQPERMHVLLVLSWSSGSGSFVQ 60
DB 14 LLLAQTFLLFLVSRP-----GSSSPAGGEDRVHVLLVLSWSSGSGSFVQ 59
QY 61 LFGCHDPDVFYLMPEPAWVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 120
DB 60 LFGCHDPDVFYLMPEPAWVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 118
QY 121 SSLFQWNSRALCSAPACDIIPDEIIPRAHCELLCSQOPFEVVEKACRSYSHVVLKEVRF 180
DB 119 SAFFNWATSRALCSPPACSAFPRGTTSKQDVCKTLCTROPFFSLAREACRSYSHVVLKEVRF 178
QY 181 FNLQSLYPLLKDPSPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 240
DB 179 FNLQVLYPLSDPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 237

Query Match
Best Local Similarity 49.5%; Score 1008; DB 9; Length 390;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 1 MLLPKMKLLFLVVSQMAILALFFHMYSHNISLSMKAQPERMHVLLVLSWSSGSGSFVQ 60
DB 14 LLLAQTFLLFLVSRP-----GSSSPAGGEDRVHVLLVLSWSSGSGSFVQ 59
QY 61 LFGCHDPDVFYLMPEPAWVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 120
DB 60 LFGCHDPDVFYLMPEPAWVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 118
QY 121 SSLFQWNSRALCSAPACDIIPDEIIPRAHCELLCSQOPFEVVEKACRSYSHVVLKEVRF 180
DB 119 SAFFNWATSRALCSPPACSAFPRGTTSKQDVCKTLCTROPFFSLAREACRSYSHVVLKEVRF 178
QY 181 FNLQSLYPLLKDPSPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 240
DB 179 FNLQVLYPLSDPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 237

```

QY 241 QPYVVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLAPVAPQATSRMYEFVGLF 298
Db 238 PHURLIREVCRSHVRIAEATKP-PPFLRGYRLVRFEDLAREPLAEIRALYAFGTJTL 296
QY 299 LPHLOTWVHNIRGKMGD--HAFHTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 356
Db 297 TPQLEAWIHNITHSGSIGKPIEAFHTSSRNAENVSOAWEHALPFTKILRVQVCAGALQL 356
QY 357 LGRVHVSQEQBNLLDLL-----STWVPE 383
Db 357 LGRYPYSADQQRDLTLVLRPGDPHFSWASPD 390

RESULT 9
US-09-927-602-8
; Sequence 8, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-8

Query Match 40.3%; Score 821; DB 9; Length 171;
Best Local Similarity 60.0%; Pred. No. 1.5e-76;
Matches 171; Conservative 0; Mismatches 0; Indels 114; Gaps 2;
QY 32 SLSMKAQPERMHVLLSWRSVGGQFQGHDPVFLMEPAHVMWTFKQSTAWMLH 91
Db 1 SLSMKAQPERMHVLLSWRSVGGQFQGHDPVFLMEPAHVMWTFK 52
QY 92 MAVRDILRAVFLCDMSVFDAYMEPGPRQSSILFQWENSALCSAPACDIIPQDEII 151
Db 53 ----- 52
QY 152 CRLLCQOPPEVVEKACRSYSHVVLKEVFFNLQSLYPLKDPNSLNLHVLVLRDPRAVF 211
Db 53 -----KACRSYSHVVLKEVFFNLQSLYPLKDPNSLNLHVLVLRDPRAVF 98
QY 212 RSRRTKGLMDSRVWGQHEOKLKEDQPYVMQVICSQLEIYKTIQSLPKALQERY 271
Db 99 RSRRTKGLMDS-----KTIQSLPKALQERY 126
QY 272 LLVRYEDLAPVAPQATSRMYEFVGLFPLHLOTWVHNIRGKMG 316
Db 127 LLVRYEDLAPVAPQATSRMYEFVGLFPLHLOTWVHNIRGKMG 171

RESULT 10
US-10-212-933-2
; Sequence 2, Application US/10212933
; Publication No. US20030008366A1
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-
; TITLE OF INVENTION: SULFOTRANSFERASE AND

; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/10/212,933
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-212-933-2

Query Match 29.4%; Score 598.5; DB 14; Length 483;
Best Local Similarity 36.8%; Pred. No. 7.3e-53;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;
QY 41 ERMHVLLSSWRSGSVFGOLFQGHDPVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRA 100
Db 116 KRQLVYVFTTWSRSGSVFGOLFQGHDPVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRA 175
QY 101 VFLCDMSVFDAYMEPGPRQSSILFQWENSALCSAPACDIIPQDEIIIPRAHCRLL- 155
Db 176 LYRCDLSVFLYSPAGSGGRNLTTLGFGAATNKVVCSSPLCPAY-RKEVVGVLVDRVCK 234
QY 156 -CSQOPPEVVEKACRSYSHVVLKEVFFNLQSLYPLKDPNSLNLHVLVLRDPRAVFRS 214
Db 235 KCPQRLAREFEBCRKRYTVVIXGVRFVAVLAPLLKDPALDKVLHVRDPRAVASSR 294
QY 215 ERKTKGLMDSRVWGQHEOKLKEDQ-----PYV---VMQVICSQ 252
Db 295 IRRSHGLIRSLQVRSRDPRAHMFLEAGHKLGAKEGMPADYHALGANEVICNS 354
QY 253 QLEIYKTIQSLPKALQERYLLVRYEDLAPVAPQATSRMYEFVGLFPLHLOTWVHNIRG 312
Db 355 MAKTLQALQPPDWLQGHVLYVRYEDLVGDFVKTLLRVYDFVGLLVSPEMEQLANMTSG 414
QY 313 KMGDHAFTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNLGVRHVSQEQBNLL 372
Db 415 SGSSSKPFVVSARNATQAANWRTALTFOQIKQVEEFCYQPMVLTGVRVNSPEEVKDL 474
QY 373 LDLL 376
Db 475 KTL 478

RESULT 11
US-10-212-933-4
; Sequence 4, Application US/10212933
; Publication No. US20030008366A1
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/10/212,933
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844


```

QY 157 SQPFVWVKACRSYSHVVLKEVRFFNLSQSLYPLKDPNSLNHIVHLVRDPRVRSRER 216
Db 186 GLNLTVAAEACRERSHVAIKTVRPEVNDLRALVEDPRLNLKVIQLVRDPRGILASRSE 245
QY 217 TKGDLMSDRISVWGQHEQKLKEDOPYV-----MQVICSQSLYIKTQSLPKALQERYL 272
Db 246 TFRDYRLWRLWYGTGR-----KPYNLDVTQLTTCDFNSVSTGLMRPPWLKGYM 298
QY 273 LVRYEDLARAPVAQTSRMVYFVGLFELPHLQTVVHNITRG-KGMGDHAFHTNARDALNVS 331
Db 299 LVRYEDLARPNMKTEETIEVGLGIPLDSDHVAWQNTNRTGDDPTLGKHKYGTVRNSAATA 357
QY 332 QAWNSLPYKVSRLQKAGCDAMNLLGYHRVSEGEORNLILLDL 376
Db 358 EKWFRFLSYDIVAFNAQCOVLAQGYKIAASEBELKNPVSLSV 402

RESULT 14
US-10-021-660-128
; Sequence 128, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-000710US
; CURRENT APPLICATION NUMBER: US/10/021,660
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US/09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-128

Query Match 26.9%; Score 549; DB 14; Length 411;
Best Local Similarity 34.3%; Pred. No. 7.8e-48;
Matches 139; Conservative 64; Mismatches 160; Indels 42; Gaps 10;

QY 6 KMKLLLLFLVS---QWAILALFFHMYSHNLSLSMKAAQPERM-----HVLV 47
Db 6 KAVLLALASIAIQYTAINTFTAKGFHTCPGLAEAGLAERLCEESPTTFAYNLKRKTHILI 65
QY 48 LSSWRSGSFVQGLFGQHPDVFYLMPEAWHVMWT-----FKQSTA----WMLHMAVRDLIR 99
Db 66 LATTRSGSFVQGLFNQHLVDVLEFYLHVQNTLIPRTQKSPADRVMLGASRDLLR 125
QY 100 AVFLCDMSVFDAYMEPGPRQSS--LFQWENSRALCSAPACDII-PODEIIPRAHCRLLC 156
Db 126 SLYDCDLYLENIYKIPPPVYHNTTDRIFRAGASRLCSRPCVCDPPGPDVLEEGDCVRKC 185
QY 157 SQPFVWVKACRSYSHVVLKEVRFFNLSQSLYPLKDPNSLNHIVHLVRDPRVRSRER 216
Db 186 GLNLTVAAEACRERSHVAIKTVRPEVNDLRALVEDPRLNLKVIQLVRDPRGILASRSE 245
QY 217 TKGDLMSDRISVWGQHEQKLKEDOPYV-----MQVICSQSLYIKTQSLPKALQERYL 272
Db 246 TFRDYRLWRLWYGTGR-----KPYNLDVTQLTTCDFNSVSTGLMRPPWLKGYM 298
QY 273 LVRYEDLARAPVAQTSRMVYFVGLFELPHLQTVVHNITRG-KGMGDHAFHTNARDALNVS 331
Db 299 LVRYEDLARPNMKTEETIEVGLGIPLDSDHVAWQNTNRTGDDPTLGKHKYGTVRNSAATA 357
QY 332 QAWNSLPYKVSRLQKAGCDAMNLLGYHRVSEGEORNLILLDL 376

```

```

Db 358 EKWFRFLSYDIVAFNAQCOVLAQGYKIAASEBELKNPVSLSV 402

RESULT 15
US-09-927-602-6
; Sequence 6, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Tomoya O.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-6

Query Match 25.5%; Score 519.5; DB 9; Length 169;
Best Local Similarity 41.1%; Pred. No. 2.5e-45;
Matches 113; Conservative 15; Mismatches 34; Indels 113; Gaps 3;

QY 42 RMHVLVLSWRSGSFVQGLFGQHPDVFYLMPEAWHVMWTFKQSTAWMLHMAVRDLIRAV 101
Db 8 RVHVLVLSWRSGSFVQGLFGQHPDVFYLMPEAWHVMWT----- 47
QY 102 FLCMSVFDAYMEPGPRQSSLFQWENSRALCSAPACDIIPODEIIPRAHCRLLCSQQPF 161
Db 48 ----- 47
QY 162 EVVEKACRSYSHVVLKEVRFFNLSQSLYPLKDPNSLNHIVHLVRDPRVRSRERTKGD 221
Db 48 --LSEACRSYSHVVLKEVRFFNLSQSLYPLKDPNSLNHIVHLVRDPRVRSRERTKAKAL 105
QY 222 MDSRIVMGQHEQKLKEDOPYVYMQVICQSLYIKTQSLPKALQERYLLVRYEDLAR 281
Db 106 ARDNEA-----TLKP-PPFLRGYRLVRPEDLAR 134
QY 282 APVAQTSRMVYFVGLFELPHLQTVVHNITRGKMG 316
Db 135 EPLAETRALYFTGLSLTPQLEAWIHNTHTGSGFG 169

Search completed: May 6, 2004, 10:57:20
Job time : 49 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:48:57 ; Search time 22 Seconds
(without alignments)
905.801 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKMKLLFLVSQMAIL.....FQNNLLDLSTWTVPQI 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/aaa/5A COMB.pcp:*
2: /cgn2_6/prodata/2/aaa/5B COMB.pcp:*
3: /cgn2_6/prodata/2/aaa/5A COMB.pcp:*
4: /cgn2_6/prodata/2/aaa/6B COMB.pcp:*
5: /cgn2_6/prodata/2/aaa/PCTUS COMB.pcp:*
6: /cgn2_6/prodata/2/aaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	100.0	386	3 US-09-045-284A-2	Sequence 2, Appli
2	2038	100.0	386	4 US-09-190-911-1	Sequence 1, Appli
3	1936	95.0	386	4 US-09-786-240-11	Sequence 11, Appl
4	598.5	29.4	483	3 US-09-263-023-2	Sequence 2, Appli
5	598.5	29.4	483	4 US-09-471-867-2	Sequence 2, Appli
6	587.5	28.8	484	3 US-09-263-023-4	Sequence 4, Appli
7	587.5	28.8	484	4 US-09-471-867-4	Sequence 4, Appli
8	549	26.9	411	4 US-09-015-188-2	Sequence 2, Appli
9	500.5	24.6	458	2 US-08-655-878-2	Sequence 2, Appli
10	482	23.7	479	2 US-08-899-514-2	Sequence 2, Appli
11	100.5	4.9	668	4 US-09-811-469-6	Sequence 6, Appli
12	94.5	4.6	848	4 US-09-540-824-27	Sequence 27, Appl
13	92	4.5	359	3 US-09-150-133-11	Sequence 11, Appl
14	92	4.5	359	3 US-09-150-141-11	Sequence 11, Appl
15	92	4.5	359	3 US-09-374-493-11	Sequence 11, Appl
16	92	4.5	359	3 US-09-374-824-11	Sequence 11, Appl
17	92	4.5	359	3 US-09-374-492-11	Sequence 11, Appl
18	92	4.5	359	4 US-09-785-343-11	Sequence 11, Appl
19	87.5	4.3	380	3 US-09-150-133-9	Sequence 9, Appli
20	87.5	4.3	380	3 US-09-150-141-9	Sequence 9, Appli
21	87.5	4.3	380	3 US-09-374-493-9	Sequence 9, Appli
22	87.5	4.3	380	3 US-09-374-824-9	Sequence 9, Appli
23	87.5	4.3	380	4 US-09-374-492-9	Sequence 9, Appli
24	87.5	4.3	380	4 US-09-785-343-9	Sequence 9, Appli
25	86.5	4.2	831	1 US-08-073-384C-5	Sequence 5, Appli
26	86.5	4.2	831	1 US-08-254-359A-5	Sequence 5, Appli
27	86.5	4.2	831	1 US-08-483-043-5	Sequence 5, Appli

28 86.5 4.2 831 1 US-08-481-238-5 Sequence 5, Appli
29 86.5 4.2 831 2 US-08-471-066B-5 Sequence 5, Appli
30 86.5 4.2 831 2 US-08-484-956-5 Sequence 5, Appli
31 86.5 4.2 831 2 US-08-757-653-5 Sequence 5, Appli
32 86.5 4.2 831 2 US-08-599-491-5 Sequence 5, Appli
33 86.5 4.2 831 2 US-08-756-386-5 Sequence 5, Appli
34 86.5 4.2 831 2 US-08-823-516-5 Sequence 5, Appli
35 86.5 4.2 831 3 US-08-882-853A-5 Sequence 5, Appli
36 86.5 4.2 831 3 US-08-759-038-5 Sequence 5, Appli
37 86.5 4.2 831 3 US-08-758-314-5 Sequence 5, Appli
38 86.5 4.2 831 4 US-09-350-309-5 Sequence 5, Appli
39 86.5 4.2 831 4 US-08-520-946-5 Sequence 5, Appli
40 86.5 4.2 831 4 US-09-684-938-5 Sequence 5, Appli
41 86.5 4.2 831 4 US-09-308-825A-5 Sequence 5, Appli
42 86.5 4.2 831 4 US-08-758-282B-5 Sequence 5, Appli
43 86.5 4.2 831 4 US-09-855-378A-5 Sequence 5, Appli
44 84.5 4.1 2549 4 US-08-265-967C-1 Sequence 1, Appli
45 84.5 4.1 2549 4 US-08-305-790B-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 100.0%; Score 2038; DB 3; Length 386;
Best Local Similarity 100.0%; Pred. No. 38-212;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 MLLPKMKLLFLVSQMAILFLPHMYSHNISLSMKAQPERMHVLVLSWRSGSFFVQ 60
Db 1 MLLPKMKLLFLVSQMAILFLPHMYSHNISLSMKAQPERMHVLVLSWRSGSFFVQ 60
Oy 61 LFGQHPDVFYLMSPAMHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPQRRQ 120
Db 61 LFGQHPDVFYLMSPAMHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPQRRQ 120
Oy 121 SSIFQWNSRALSACACDIIPODEIIPRAHCHLLCSQOPFEVVEKACRSYSHVLEKVR 180
Db 121 SSIFQWNSRALSACACDIIPODEIIPRAHCHLLCSQOPFEVVEKACRSYSHVLEKVR 180
Oy 181 FFNLQSLYPLKDPSPNLHIVHLVRDPRAVFRSRTKGLMIDSRIVMGQHEQKLKED 240
Db 181 FFNLQSLYPLKDPSPNLHIVHLVRDPRAVFRSRTKGLMIDSRIVMGQHEQKLKED 240
Oy 241 QPYVMQVICSQLEIYKTIQSIPKALQERYLLVRVEDLARAPVAQTSRMRYEVEGLEFLP 300
Db 241 QPYVMQVICSQLEIYKTIQSIPKALQERYLLVRVEDLARAPVAQTSRMRYEVEGLEFLP 300
Oy 301 HLQTVHNIITRGKMGDHAFTNARDALNVSQARWSLPYEKYSRLQKACGDAMNLLGYR 360
Db 301 HLQTVHNIITRGKMGDHAFTNARDALNVSQARWSLPYEKYSRLQKACGDAMNLLGYR 360
Oy 361 HVRSQQRNLLDLSTWTVPQI 386

Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

RESULT 2

US-09-190-911-1
 ; Sequence 1, Application US/09190911
 ; Patent No. 6365365
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/09/190,911
 ; EARLIER FILING DATE: 1998-11-12
 ; EARLIER FILING DATE: 1998-03-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 ; US-09-190-911-1

Query Match 100.0%; Score 2038; DB 4; Length 386;
 Best Local Similarity 100.0%; Pred. No. 3e-212;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSFVQ 60
 Db 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSFVQ 60
 Qy 61 LFGQHPDVFYLMPEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 Db 61 LFGQHPDVFYLMPEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 Qy 121 SSLFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQQPFVEVEKACRSYSHVYLKEVR 180
 Db 121 SSLFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQQPFVEVEKACRSYSHVYLKEVR 180
 Qy 181 FNLQSLYPLKDPKPSLNHLVHVRDPAVFRSRTKGLMIDSRIVMGQHEQKLKED 240
 Db 181 FNLQSLYPLKDPKPSLNHLVHVRDPAVFRSRTKGLMIDSRIVMGQHEQKLKED 240
 Qy 241 QPYVVMQVICQSOLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 Db 241 QPYVVMQVICQSOLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 Qy 301 HLQTVHNTTRGKMGDGHAFHTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
 Db 301 HLQTVHNTTRGKMGDGHAFHTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
 Qy 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
 Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

RESULT 3

US-09-786-240-11
 ; Sequence 11, Application US/09786240
 ; Patent No. 6558935
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE PHARMACEUTICALS, INC.
 ; APPLICANT: TANG, Y. Tom
 ; APPLICANT: CORLEY, Neil C.
 ; APPLICANT: GUEGLER, Karl J.
 ; APPLICANT: BAUGHN, Mariah R.
 ; APPLICANT: LAL, Preeti
 ; APPLICANT: YUE, Henry
 ; APPLICANT: HILLMAN, Jennifer L.
 ; APPLICANT: AZINZAI, Yalda

; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
 ; FILE REFERENCE: PF-0592 PCT
 ; CURRENT APPLICATION NUMBER: US/09/786,240
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
 ; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1999-05-11
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 11
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
 ; US-09-786-240-11

Query Match 95.0%; Score 1936; DB 4; Length 386;
 Best Local Similarity 95.6%; Pred. No. 3.4e-201;
 Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSFVQ 60
 Db 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSFVQ 60
 Qy 61 LFGQHPDVFYLMPEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 Db 61 LFGQHPDVFYLMPEPAWHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 Qy 121 SSLFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQQPFVEVEKACRSYSHVYLKEVR 180
 Db 121 SSLFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQQPFVEVEKACRSYSHVYLKEVR 180
 Qy 181 FNLQSLYPLKDPKPSLNHLVHVRDPAVFRSRTKGLMIDSRIVMGQHEQKLKED 240
 Db 181 FNLQSLYPLKDPKPSLNHLVHVRDPAVFRSRTKGLMIDSRIVMGQHEQKLKED 240
 Qy 241 QPYVVMQVICQSOLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 Db 241 QPYVVMQVICQSOLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 Qy 301 HLQTVHNTTRGKMGDGHAFHTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
 Db 301 HLQTVHNTTRGKMGDGHAFHTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
 Qy 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
 Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

RESULT 4

US-09-263-023-2
 ; Sequence 2, Application US/09263023
 ; Patent No. 6037159
 ; GENERAL INFORMATION:
 ; APPLICANT: Uchimura, Kenji
 ; APPLICANT: Muramatsu, Hideki
 ; APPLICANT: Kadomatsu, Kenji
 ; APPLICANT: Kannagi, Re-ji
 ; APPLICANT: Habuchi, Osami
 ; APPLICANT: Muramatsu, Takashi
 ; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
 ; FILE REFERENCE: TOYAMA1.001AUS
 ; CURRENT APPLICATION NUMBER: US/09/263,023
 ; EARLIER FILING DATE: 1999-03-05
 ; EARLIER APPLICATION NUMBER: JP 10-54007
 ; EARLIER FILING DATE: 1998-03-05
 ; EARLIER APPLICATION NUMBER: JP 10-177844
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2

```
; LENGTH: 483
; TYPE: PR1
; ORGANISM: Mus musculus
US-09-263-023-2

Query Match
Best Local Similarity 29.4%; Score 598.5; DB 3; Length 483;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWRSRGSSFFVGLFQHPDVFVLMPEPAHVMWTFKQSTAMMLHMAVRLIRA 100
DB 116 KROLVYVFTTWRSRGSSFFGELFNQNEVFVLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 175
QY 101 VFLCDMSVFDAYMEPPGRRQS-----SLFQWENSRALCSAPACDIIPODEIIPRAHCRLL- 155
DB 176 LYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDRVCK 234
QY 156 -CSQQPEVVEKACRSYSHVVLKEVPFNLSYPLLKDPKPSLNHLHVLVRDPAVFRSR 214
DB 235 KCPQRLARFEERCRTYTVIKGRVDFVAVLAPLKDPAFLDLKVIHLVRDPAVASSR 294
QY 215 ERTKGDLMIDSRIVM-----GQHEQKLKEDQ-----PYV---VMQVICS 252
DB 295 IRRSHGLIRSLQVVRSDPRAHRMPEFLAAGHKLGAKEGMPGADYHALGAMEVICNS 354
QY 253 QLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPLHLQTVWVHNITRG 312
DB 355 MAKTLQALQPPDWLQGHVLYVRYEDLVGDPVKTLLRVYDFVGLLVSPENQFALNMTSG 414
QY 313 KGMGDHAFHTNARDALNVSAQWWSLPEYKVSRLQKACGDAMNLLGYRHSRSEQORNL 372
DB 415 SGSSSKPFVVSARNATQANAWRTALTFTQIKQVEEFYQPMVAVLGYRVNSPEEVKDL 474
QY 373 LDLL 376
DB 475 KTL 478

RESULT 5
US-09-471-867-2
; Sequence 2, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; EARLIER FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PR1
; ORGANISM: Mus musculus
US-09-471-867-2

Query Match
Best Local Similarity 29.4%; Score 598.5; DB 4; Length 483;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWRSRGSSFFVGLFQHPDVFVLMPEPAHVMWTFKQSTAMMLHMAVRLIRA 100
```

```
DB 116 KROLVYVFTTWRSRGSSFFGELFNQNEVFVLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 175
QY 101 VFLCDMSVFDAYMEPPGRRQS-----SLFQWENSRALCSAPACDIIPODEIIPRAHCRLL- 155
DB 176 LYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDRVCK 234
QY 156 -CSQQPEVVEKACRSYSHVVLKEVPFNLSYPLLKDPKPSLNHLHVLVRDPAVFRSR 214
DB 235 KCPQRLARFEERCRTYTVIKGRVDFVAVLAPLKDPAFLDLKVIHLVRDPAVASSR 294
QY 215 ERTKGDLMIDSRIVM-----GQHEQKLKEDQ-----PYV---VMQVICS 252
DB 295 IRRSHGLIRSLQVVRSDPRAHRMPEFLAAGHKLGAKEGMPGADYHALGAMEVICNS 354
QY 253 QLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPLHLQTVWVHNITRG 312
DB 355 MAKTLQALQPPDWLQGHVLYVRYEDLVGDPVKTLLRVYDFVGLLVSPENQFALNMTSG 414
QY 313 KGMGDHAFHTNARDALNVSAQWWSLPEYKVSRLQKACGDAMNLLGYRHSRSEQORNL 372
DB 415 SGSSSKPFVVSARNATQANAWRTALTFTQIKQVEEFYQPMVAVLGYRVNSPEEVKDL 474
QY 373 LDLL 376
DB 475 KTL 478

RESULT 6
US-09-263-023-4
; Sequence 4, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-263-023-4

Query Match
Best Local Similarity 28.8%; Score 587.5; DB 3; Length 484;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWRSRGSSFFVGLFQHPDVFVLMPEPAHVMWTFKQSTAMMLHMAVRLIRA 100
DB 117 KRWVYVFTTWRSRGSSFFGELFNQNEVFVLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 176
QY 101 VFLCDMSVFDAYMEPPGRRQS-----SLFQWENSRALCSAPACDIIPODEIIPRAHCRLL- 155
DB 177 LYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDRVCK 235
QY 156 -CSQQPEVVEKACRSYSHVVLKEVPFNLSYPLLKDPKPSLNHLHVLVRDPAVFRSR 214
DB 236 KCPQRLARFEERCRTYTVIKGRVDFVAVLAPLKDPAFLDLKVIHLVRDPAVASSR 295
QY 215 ERTKGDLMIDSRIVM-----GQHEQKLKEDQ-----DQPIY---VMQVICS 252
```

```
Db 296 ISRHGLRESQVVRSDPRRAHVPFLEAAGHKLGAKEGVGGPADYHALGAMEVICNS 355
QY 253 QLEIYKTTQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELPHLQTVWHNITRG 312
Db 356 MAKTLQALQPDWLGQHYLVVRYEDLVGDPVKTLRRVYDFVGLVSPMEQFALNWTSG 415
QY 313 KMGDGHAFHTNARDALNVSQAWWSLPEYKVSRLQKACGDAMNLLGYRHVSEQEQNLL 372
Db 416 SGSSSKPFVVSARNATQANAWRTALTTCQIKQVEEFCYQPMVGLYGVNSPBEVKDLS 475
QY 373 LDLL 376
Db 476 KTL 479

RESULT 7
US-09-471-867-4
; Sequence 4, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kanragi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 28.8%; Score 587.5; DB 4; Length 484;
Best Local Similarity 36.0%; Pred. No. 6.8e-55;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWSGSSFGQLFGQHPDVPYLMPEPAWHVMTFKOSTAMLMHVAVRDLIRA 100
Db 117 KHMVMYVTTWRSGSSFGELFNQNPVEFFLYEPVWHVMQKLYPCDAVSLQGAARDLSA 176
QY 101 VFLCDMSVFDAYMEGPPRRQS----SLQWENSRLCSAPACDIIPDEIIPRAHCRLL- 155
Db 177 LYRCDLSVFLYSPAGSGGRNLTIGIFGAATNKVCSPLCPAY-RKEVGLVDDRVCK 235
QY 156 -CSQOPFEVVEKACRSYSHVVKVRFNPNLSPLPKDPSLNHLHVLHVRDPAVFRSR 214
Db 236 KCPQRLARFEECKRYRTLKIGVRVFDVAVLPLRDLRDPALDLKVLHLVRDPAVASSR 295
QY 215 ERTKGLMIDSRIVN-----GHEQKLKE---DQVY---VMQVICSQ 252
Db 296 ISRHGLRESQVVRSDPRRAHVPFLEAAGHKLGAKEGVGGPADYHALGAMEVICNS 355
QY 253 QLEIYKTTQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELPHLQTVWHNITRG 312
Db 356 MAKTLQALQPDWLGQHYLVVRYEDLVGDPVKTLRRVYDFVGLVSPMEQFALNWTSG 415
QY 313 KMGDGHAFHTNARDALNVSQAWWSLPEYKVSRLQKACGDAMNLLGYRHVSEQEQNLL 372
Db 416 SGSSSKPFVVSARNATQANAWRTALTTCQIKQVEEFCYQPMVGLYGVNSPBEVKDLS 475
```

```
QY 373 LDLL 376
Db 476 KTL 479
```

RESULT 8

```
US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2
```

```
Query Match 26.9%; Score 549; DB 4; Length 411;
Best Local Similarity 34.3%; Pred. No. 7.9e-51;
Matches 139; Conservative 64; Mismatches 160; Indels 42; Gaps 10;
```

```
QY 6 KMKLLFLVS---QMAILALFFHMYSHNISSLSMAKQPERM-----HVLV 47
Db 6 KAVLLIALASIAIQVTAIRTTAKSFHTCPGLAEAGLAERLCESTPAYNLSRKHILI 65
QY 48 LSSWSGSSFGQLFGQHPDVPYLMPEPAWHVMT---FKQSTA---WMLHMAVRDLIR 99
Db 66 LATTSGSSFGQLFNQHLVDVILFELPYHVQNTLIPRTQCKSPADRVMLGASRDLLR 125
QY 100 AVFLCDMSVFDAYMEGPPRRQS--LQWENSRLCSAPACDIIPDEIIPRAHCRLLC 156
Db 126 SLYDCDLYFLENYKIPPPVNVHTTDRFRGASRLCSRPVCDPPGPADLVLEEGDCVRKC 185
QY 157 SQOPFEVVEKACRSYSHVVKVRFNPNLSPLPKDPSLNHLHVLHVRDPAVFRSRER 216
Db 186 GLNLTVAAEACRERSHVAIKTVRFVNDLGRALVEDPRLNKLVIQLVRDPRGIIASRSE 245
QY 217 TKGDLMIDSRIVMGQHEQKLKEDQPYVY---MQVICQSLIYKTIQSLPKALQERYL 272
Db 246 TFRDLYRLWRLWVGTCR-----KPYNLDVTLTTCEDFNSVSTGLMRPPLKGYM 298
QY 273 LVRYEDLARAPVAQTSRMVYFVGLFELPHLQTVWHNITRG-KMGDGHAFHTNARDALNV 331
Db 299 LVRYEDLARPNPKKTEIEIYGLGIPLDGSHVARWICQNTNRGDTPLGKHKYGTVRNSAATA 357
QY 332 QAWWSLPEYKVSRLQKACGDAMNLLGYRHVSEQEQNLLDLL 376
Db 358 EKWRFSLVDIVAFQAQACQVLAQLGYKIAASEELKNPVSIV 402
```

RESULT 9

```
US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASAKAZU
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
```

COUNTRY: ;
 ZIP: ;
 COMPUTER READABLE FORM: ;
 MEDIUM TYPE: Floppy disk ;
 COMPUTER: IBM PC compatible ;
 OPERATING SYSTEM: PC-DOS/MS-DOS ;
 SOFTWARE: PatentIn ;
 CURRENT APPLICATION DATA: ;
 APPLICATION NUMBER: US/08/655,878 ;
 FILING DATE: ;
 CLASSIFICATION: 435 ;
 PRIOR APPLICATION DATA: ;
 APPLICATION NUMBER: ;
 FILING DATE: ;
 ATTORNEY/AGENT INFORMATION: ;
 NAME: ;
 REGISTRATION NUMBER: ;
 REFERENCE/DOCKET NUMBER: ;
 TELEPHONE: ;
 TELEFAX: ;
 INFORMATION FOR SEQ ID NO: 2: ;
 SEQUENCE CHARACTERISTICS: ;
 LENGTH: 458 ;
 TYPE: amino acid ;
 TOPOLOGY: linear ;
 MOLECULE TYPE: protein ;
 US-08-655-878-2 ;

Query Match 24.6%; Score 500.5; DB 2; Length 458;
 Best Local Similarity 33.2%; Pred. No. 1.7e-45;
 Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;
 QY 34 LNKAKQPERMHLVLSWSRSGSFVQQLFGQHPDVFLMEPAWHV--WMTFKQSTAMLMH 91
 DB 104 LGIAAEPRRHVLLMATTGSGFGEFFNQGNIFLPEPLWHIERTVTPEPGGANAVG 163
 QY 92 MAV--RDLIRAVFLCMSVFDAYMEPPRRQ--SSLFQWNSRALCSAPACDIIIQ-DEI 146
 DB 164 SALVYRDVQLCLLCOLYILESFISPAPEHLTAALFRGSGSHSCEEPVC--TPSLKKV 221
 QY 147 IPRACR-LICSQPPEVVEKACRSYSHVVLKEVRFNQLSLYPLKDPSLMLHLVHLVR 205
 DB 222 FEKYCHKNRCGPLNTLAAEACRRKQHAKLVIRIQLEFLQPLAEDPRDLRIIQLVR 281
 QY 206 DPRVRSRERTKGLMIDSRIVMGQHEQKLK-----KEDQPYVMQVICS-QLE 255
 DB 282 DPRVLVSR-----MV---AFSGKYESKWKWAAEGEAPLOEDE-VQSLRGNCESIRLS 330
 QY 256 IYKTIQSLPKALQERYLLVRYEDLAPVPAQTSRMVEFVGLFELPHLOTWVHNITRGKM 315
 DB 331 AELGLRQ-FRWLGRVLYRYEDVARAPLRKALEMYRFAGIHPTQVEWIRANTQAP-Q 388
 QY 316 GDHAFHTNARDALNVSQAWNSLPYKVSRLQKACGDAMNLLGYRHVRSEQEQRNLLLDL 375
 DB 389 DSGNIYSTOKNSSEQEKWRFSFPFKLAQVVDACEPAMRLFGYKLASSAQELTNRSLSL 448
 QY 376 L 376
 DB 449 L 449

RESULT 10
 US-08-899-514-2
 ; Sequence 2, Application US/08899514
 ; Patent No. 5910581
 ; GENERAL INFORMATION:
 ; APPLICANT: HABUCHI, OSAMI
 ; APPLICANT: FUKUTA, WASAKAZU
 ; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 ; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 ; TITLE OF INVENTION: FOR THE POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS: ;
 ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP ;
 STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR ;
 CITY: NEWPORT BEACH ;
 STATE: CALIFORNIA ;
 COUNTRY: US ;
 ZIP: 92660 ;
 COMPUTER READABLE FORM: ;
 MEDIUM TYPE: Floppy disk ;
 COMPUTER: IBM PC compatible ;
 OPERATING SYSTEM: PC-DOS/MS-DOS ;
 SOFTWARE: PatentIn ;
 CURRENT APPLICATION DATA: ;
 APPLICATION NUMBER: US/08/899,514 ;
 FILING DATE: ;
 CLASSIFICATION: 435 ;
 PRIOR APPLICATION DATA: ;
 APPLICATION NUMBER: ;
 FILING DATE: ;
 ATTORNEY/AGENT INFORMATION: ;
 NAME: DANIEL E. ALTMAN ;
 REGISTRATION NUMBER: 34,115 ;
 REFERENCE/DOCKET NUMBER: TOYAM21.001AUS ;
 TELECOMMUNICATION INFORMATION: ;
 TELEPHONE: 714 760 0404 ;
 TELEFAX: 714 760 9502 ;
 INFORMATION FOR SEQ ID NO: 2: ;
 SEQUENCE CHARACTERISTICS: ;
 LENGTH: 479 ;
 TYPE: amino acid ;
 TOPOLOGY: linear ;
 MOLECULE TYPE: protein ;
 US-08-899-514-2 ;

Query Match 23.7%; Score 482; DB 2; Length 479;
 Best Local Similarity 32.4%; Pred. No. 1.8e-43;
 Matches 113; Conservative 71; Mismatches 141; Indels 24; Gaps 10;
 QY 42 RMHVLVLSWSRSGSFVQQLFGQHPDVFLMEPAWHVMT--FKQSTAMLMHMAV--RDL 97
 DB 132 RRRHLLMATTGSGFGEFFNQGNIFLPEPLWHIERTVSPFGGANAAGSALVYRDV 191
 QY 98 IRVAVFLCMSVFDAYMEPPRRQSSLFQWE--NSRALCSAPACDIIIPQDEIIIPRAHCR-L 154
 DB 192 LKQLFLCLVLYLHFTPLPEDHLTQFMFREGSSRSICEDPVCTPFVK-KVFEKYHCKNR 250
 QY 155 LCSQPPPEVVEKACRSYSHVVLKEVRFNQLSLYPLKDPSLMLHLVHLVDRPRAVRSR 214
 DB 251 RCGPLNVTLAAEACRRKKEHAKLVIRIQLEFLQPLAEDPRDLRVQLVDRPRAVLASR 310
 QY 215 -----ERTKGLMIDSRIVMGQHEQKLKEDQPYVMQVICS-QLEIYKTIQSLPKAL 267
 DB 311 MVAFAQYKTKWKWLDDEGGDGLREEVQR-----LRGNCESIRLSAELGRQ-PAWL 362
 QY 268 QERYLLVRYEDLAPVPAQTSRMVEFVGLFELPHLOTWVHNITRGKMGDHAFTHARDA 327
 DB 363 RGRVLYRYEDVARGPLQKAREMYFPAGIPLTQVEDWIKNTQAAHDSGGIYSTQ-KNS 421
 QY 328 LNVSQAWNSLPYKVSRLQKACGDAMNLLGYRHVRSEQEQRNLLLDL 376
 DB 422 SEQEKWRFSMPFKLAQVVDACEPAMRLFGYKLARDAAALTNRSVSL 470

RESULT 11
 US-08-811-469-6
 ; Sequence 6, Application US/09811469
 ; Patent No. 6551809
 ; GENERAL INFORMATION:
 ; APPLICANT: YAN, Chunhua et al.
 ; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
 ; TITLE OF INVENTION: AND USES THEREOF
 ; FILE REFERENCE: CJ001171

Search completed: May 6, 2004, 10:52:32
Job time : 24 secs

```

108 VFDAYMEPGPRRQSSLFQWENSF
    |||
144 VEDDAYV-----

```

[illegible]

```

RESULT 15
US-09-374-493-11
  / Sequence 11, Application US/0937443
  / Patent No. 6204015
  / GENERAL INFORMATION:
  / APPLICANT: The Board of Regents of
  / TITLE OF INVENTION: TVROSYLPROTEIN
  / FILE REFERENCE: 5820,546
  / CURRENT APPLICATION NUMBER: US/09
  / CURRENT FILING DATE: 1999-08-13
  / EARLIER APPLICATION NUMBER: 09/15
  / EARLIER FILING DATE: 1998-09-09
  / EARLIER APPLICATION NUMBER: 60/07
  / EARLIER FILING DATE: 1998-01-29
  / EARLIER APPLICATION NUMBER: PCT/U
  / EARLIER FILING DATE: 1999-07-23
  / NUMBER OF SEQ ID NOS: 45
  / SOFTWARE: Wordperfect 8.0 (saved
  / SEQ ID NO 11
  / LENGTH: 359
  / TYPE: PRT
  / ORGANISM: Caenorhabditis elegans
US-09-374-493-11

```

Query Match	4.5%; SC
Best Local Similarity	19.5%; Pr
Matches	63; Conservative 48;
Qy	52 RSGSFFVGOLFQGHDPV---F---F
Db	95 RSGTTLMAILDAHPDVRGCGE
Qy	108 VFQAYMEPGPRQSLLFQWENS
Db	144 VFDDAV-----
Qy	167 ACSRSYSHWLKVEVPFNQLSYL
Db	173 ----YTAALPTIR----RLV